DOCUMENT RESUME

ED 438 325 TM 030 642

AUTHOR	Dings, Jonathan; Gong, Brian; Kingston, Neal
TITLE	KIRIS Accountability Cycle I Technical Manual. Based on the Analysis of Data from the 1991-92 through 1993-94 School Years.
INSTITUTION	Kentucky State Dept. of Education, Frankfort.
PUB DATE	1995-07-10
NOTE	350p.; For the Cycle 2 Technical Manual, see TM 030 643.
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Descriptive (141)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; *Accountability; *Educational Assessment; Educational Change; Elementary Secondary Education; *Psychometrics; Scoring; State Programs; *State Standards; Test Construction; Test Use; Testing Problems; *Testing Programs; Validity
IDENTIFIERS	Kentucky; *Kentucky Instructional Results Information System; Test Specifications

ABSTRACT

This technical manual provides all major existing information about the technical characteristics of the Kentucky Instructional Results Information System (KIRIS). The assessment component of the Kentucky Education Reform Act Accountability System was in use between 1991-92 and 1993-94 and was referred to as Cycle One. Two sets of test specifications were used to develop the KIRIS assessments. A subset of the 57 academic expectations that explicate four of Kentucky's six learner goals forms one set of the specifications, and a modification of the content framework of the National Assessment of Educational Progress forms the other. Sections of the manual, which traces the development of the tests and discusses their psychometric characteristics, are: (1) "Context"; (2) "Test Development and Item Analysis"; (3) "Administration, Scoring, and Reporting"; (4) "Generalizability"; and (5) "Appendices." Eight appendixes contain supplemental information and some forms used in the KIRIS. (Contains 175 tables and 6 figures.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 438 325

KIRIS Accountability Cycle I Technical Manual

Based on the Analysis of Data
from the 1991-92 through 1993-94
School Years

BEST COPY AVAILABLE

Produced by the Kentucky Department of Education
July 10, 1995

TM030642

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL IN MICROFICHE
AND IN ELECTRONIC MEDIA FOR ERIC
COLLECTION SUBSCRIBERS ONLY
HAS BEEN GRANTED BY

A. Arrastia

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

DEDICATION

As is all of school reform in Kentucky, this manual is dedicated to our children. May their dreams never be fettered by a lack of learning.

ACKNOWLEDGEMENTS

This manual has benefited from the work and advice of many individuals. While we are grateful for this assistance, sole responsibility for the accuracy and completeness of this manual rests with the authors. Table A presents an alphabetical list of those involved with this manual and their affiliations.

DESIGN. The primary designers of this manual are Jonathan Dings, Sean Donovan, Neal Kingston, and Ed Reidy.

AUTHORSHIP. The primary authors of this manual are Jonathan Dings, Brian Gong, and Neal Kingston. David Green co-authored the chapter on test development. This manual was based substantially on two earlier Kentucky Instructional Results Information System (KIRIS) Technical Manuals produced by Advanced Systems in Measurement and Evaluation (ASME), which were authored by Charles DePascale, Judith Gerstenblatt, Richard Hill, Stuart Kahl, and Amy Sosman.

DATA ANALYSIS. Data analyses and tables for this manual were produced by Julie Ashworth, Charles DePascale, Jonathan Dings, Sean Donovan, Brian Gong, Richard Hill, Neal Kingston, and Cecilia Hegamin-Younger.

ADVICE AND REVIEW. Throughout the first accountability cycle, Richard Hill regularly designed and carried out psychometric analyses critical to the implementation of the testing and accountability programs. Scott Trimble provided significant input on all aspects of the testing and accountability programs. On an ongoing basis, we received psychometric advice from several members of our technical advisory committee – Edward Haertel, Skip Kifer, and David Wiley. Advice related to the generalizability analyses in this manual were provided by Sheila Barron, Robert Brennan, Ed Haertel, and Robert Mislevy. Within their report, the authors of the Kentucky Institute for Educational Research (KIER) evaluation report – Mark Fenster, Daniel Stufflebeam, and Anthony Nitko – provided advice on a number of topics. We benefited also from questions and data requests from Dan Koretz, Richard Jaeger, and Robert Linn, who participated in a review of KIRIS funded by Kentucky's Office of Educational Accountability.

Jonathan Dings, Betty Edwards, Lisa Ehrlich, Brian Gong, Richard Hill, Neal Kingston, Ed Reidy, Amy Sosman, and Scott Trimble served as technical reviewers. Ann Bartosh, Amy Buren, Jean Cain, Nijel Clayton, Agnes Durbin, Mary Alice Harmon, Tricia Kerr, Jo Ann Mosier, Rhonda Sims, Sheila Vice, Kay Vincent, Kaye Warner, and Howard White reviewed the test development sections. Armando Arrastia and Judy Phillips edited the manual.

OTHER SUPPORT. Liz Brown, Teri Eads, Tina Logan, and Mary Kinsella provided secretarial support.

ABOVE ALL. Above all, we acknowledge the support and constructive criticism of many administrators, district assessment coordinators, legislators, parents, superintendents, and teachers.

Table A
People Acknowledged and Their Affiliations

Julie Ashworth, Research Assistant/Data Analyst, Allegiant Management Corporation
Ann Bartosh, Mathematics Consultant, KDE
Sheila Barron, Urban Institute
Robert Brennan, E. F Lindquist Professor of Educational Measurement, University of Iowa
Liz Brown, Olsten Temporary Services
Amy Buren, Writing Portfolio Consultant, ASME
Jean Cain, Practical Living and Vocational Studies Consultant, KDE
Nijel Clayton, Social Studies Consultant, KDE
Charles DePascale, Supervisor of Data Analysis and Scoring, ASME
Jonathan Dings, Measurement Statistician, Allegiant Management Corporation
Sean Donovan, Measurement Statistician, Allegiant Management Corporation
Agnes Durbin, Branch Manager, Curriculum and Assessment Development, KDE
Betty Edwards, Director of Curriculum and Assessment Development, KDE
Teri Eads, Secretary, KDE
Lisa Ehrlich, Kentucky Project Director, ASME
Mark Fenster, Western Michigan University
Judith Gerstenblatt, ASME
Brian Gong, Research Consultant, KDE
David Green, Curriculum and Assessment Consultant, KDE
Edward Haertel, Professor of Educational Measurement, Stanford University
Mary Alice Harmon, Writing Portfolio Consultant, KDE
Cecilia Hegamin-Younger, Graduate Student, University of Iowa
Richard Hill, President, ASME
Richard Jaeger, Professor of Educational Measurement, University of North Carolina, Greensboro
Dan McCaffrey, RAND Corporation
Stuart Kahl, Vice President, ASME
Tricia Kerr, Science Consultant, KDE
Skip Kifer, Professor of Educational Measurement, University of Kentucky
Neal Kingston, Associate Commissioner for Curriculum, Assessment, and Accountability, KDE
Mary Kinsella, Secretary, KDE
Dan Koretz, Urban Institute
Robert Linn, Professor of Educational Measurement, University of Colorado
Tina Logan, Secretary, KDE
Robert Mislevy, Distinguished Research Scientist, Educational Testing Service
Jo Ann Mosier, Manager of Mathematics Portfolios, KDE
Anthony Nitko, Professor of Educational Measurement, University of Pittsburgh
Arthur Patterson, Arts and Humanities Consultant, KDE
Judy Phillips, Assessment Implementation Consultant, KDE
Ed Reidy, Deputy Commissioner, Learning Results Services, KDE
Rhonda Sims, Social Studies Consultant, KDE
Amy Sosman, Associate Kentucky Project Director, ASME
Daniel Stufflebeam, Professor of Educational Evaluation, Western Michigan University
Sheila Vice, Branch Manager, Curriculum and Assessment Development, KDE
Kay Vincent, Branch Manager, Curriculum and Assessment Development, KDE
Kaye Warner, Language Arts Consultant, KDE
Howard White, Science Consultant, KDE
David Wiley, Dean, Northwestern University

CONTENTS

Dedication i
Acknowledgements ii
Contents iv

SECTION I: CONTEXT 1
1. Introduction 2
2. Background 3

SECTION II: TEST DEVELOPMENT AND ITEM ANALYSIS 12
3. Test Specifications 13
4. Test Development Process 54
5. Standard Setting 63
6. Item Analysis 83

SECTION III: ADMINISTRATION, SCORING, AND REPORTING 105
7. Test Administration 106
8. Scoring 119
9. Equating 124
10. Reporting to Schools and Districts 144
11. Statewide Summary of Results 147
12. Special Portfolio Issues 151

SECTION IV: GENERALIZABILITY 195
13. Reliability 196
14. Validity 199
15. Consequential Validity 215
16. Subgroup Performance 236

SECTION V: APPENDICES 243

TABLE OF TABLES

Table 2-1: Kentucky's Six Learner Goals . . . 4
Table 2-2: KRS 158.645 - Capacities Required of Students in Public Education . . . 5
Table 2-3: Academic Expectations for Goal 1: Communication Skills . . . 7
Table 2-4: Academic Expectations for Goal 2: Core Academic Concepts . . . 8
Table 2-5: Academic Expectations for Goal 5: Think and Solve Problems . . . 9
Table 2-6: Academic Expectations for Goal 6: Integrate Knowledge . . . 9
Table 2-7: Nationally Recognized Experts Who Assessed the Proposals . . . 9
Table 2-8: Kentucky Technical Review Committee . . . 10
Table 2-9: Bidders and Judgements about Their Proposals . . . 10
Table 3-1: Kentucky's Four Learner Goals that are Measured in KIRIS . . . 13
Table 3-2: Assessment of Kentucky's Academic Expectations - Goal 1 Open response Tests . . . 15
Table 3-3: Assessment of Kentucky's Academic Expectations - Goal 2 Open Response Tests . . . 16
Table 3-4: Assessment of Kentucky's Academic Expectations - Goal 5 Open Response Tests . . . 18
Table 3-5: Assessment of Kentucky's Academic Expectations - Goal 6 Open Response Tests . . . 18
Table 3-6: Assessment of Kentucky's Academic Expectations - Goal 1 Performance Events . . . 19
Table 3-7: Assessment of Kentucky's Academic Expectations - Goal 2 Performance Events . . . 20
Table 3-8: Assessment of Kentucky's Academic Expectations - Goal 2 Performance Events . . . 21
Table 3-9: Assessment of Kentucky's Academic Expectations - Goal 5 Performance Events . . . 22
Table 3-10: Assessment of Kentucky's Academic Expectations - Goal 6 Performance Events . . . 22
Table 3-11: Distribution of Items Across Academic Expectations - Grade 4 Open-Response . . . 23
Table 3-12: Distribution of Items Across Academic Expectations - Grade 8 Open-Response . . . 24
Table 3-13: Distribution of Items Across Academic Expectations - Grade 12 Open-Response . . . 25
Table 3-14: Distribution of Items Across Academic Expectations - Grade 4 Multiple-Choice . . . 26
Table 3-15: Distribution of Items Across Academic Expectations - Grade 8 Multiple-Choice . . . 27
Table 3-16: Distribution of Items Across Academic Expectations - Grade 12 Multiple-Choice . . . 28
Table 3-17: KIRIS Reading Content Frame Work . . . 30
Table 3-18: Passage Types Used in the KIRIS Reading Assessment . . . 31
Table 3-19: KIRIS Mathematics Content Framework . . . 33
Table 3-20: Topics Addressed in each Mathematics Contest Category . . . 34
Table 3-21: KIRIS Social Studies Content Framework . . . 35
Table 3-22: KIRIS Science Content Framework . . . 36
Table 3-23: KIRIS Writing Content Requirements . . . 37
Table 3-24: KIRIS Arts and Humanities Content Framework . . . 28
Table 3-25: Additional Information About the Content of the Arts and Humanities Test . . . 29
Table 3-26: KIRIS Practical Living/Vocational Studies Content Framework . . . 40
Table 3-27: Example of a Performance Event Item . . . 43
Table 3-28: Example of an Open-Response Item, With Scoring Guide . . . 45

Table 3-29: Annotated Response to Open-Response Question 46
Table 3-30: Example of Multiple-Choice Item 48
Table 3-31: Tests Administered in Grade 4 in 1991-92, 1992-93, and 1993-94 49
Table 3-32: Tests Administered in Grade 8 in 1991-92, 1992-93, and 1993-94 50
Table 3-33: Tests Administered in Grade 12 in 1992-92, 1992-93, and 1993-94 51
Table 3-34: Interdisciplinary Open-Response Items Administered in 1992-93 and 1993-94 52
Table 3-35: Numbers of KIRIS Open-Response and Multiple-Choice Questions for Each subject Per Student By Year 53
Table 4-1: Test Development Committees - Member Experience 54
Table 4-2: Test Development Committees - Member Demographics 55
Table 4-3: Major Steps in the Test Development Process 56
Table 4-4: Analyses Performed on Open-Response Items 57
Table 4-5: Sample Open-Response Item Analysis Output 58
Table 5-1: Standard Setting Committees - Member Experience 67
Table 5-2: Standard Setting Committees - Member Demographics 67
Table 5-3: General Item Scoring Guide 68
Table 5-4: Consistency of Committee Member Standard-Setting Judgements, Grade 12 Reading . . . 69
Table 5-5: Performance Standards for 1991-92 Mathematics Open-Response Tests 70
Table 5-6: Performance Standards for 1991-92 Reading Open-Response Tests 70
Table 5-7: Performance Standards for 1991-92 Science Open-Response Tests 71
Table 5-8: Performance Standards for 1991-92 Social Studies Open-Response Tests 71
Table 5-9: Performance Standards for 1992-93 Practical Living/Vocational Studies Open-Response Tests 75
Table 5-10: Performance Standards for 1992-93 Arts and Humanities Open-Response Tests 75
Table 5-11: Conversion of Writing Portfolio Score Points to Performance Levels 78
Table 5-12: Comparison of KIRIS and NAEP Standards Percent of 1991-92 Kentucky Students Scoring at Each Performance Level 82
Table 6-1: Distribution of Item Scores for Open-Response Items, Grade 4 84
Table 6-2: Distribution of Item-Total Correlations for Open-Response Items, Grade 4 85
Table 6-3: Novice VS. Apprentice, Proficient, Distinguished Distribution of Biserial Correlations for Open-Response Items, Grade 4 86
Table 6-4: Novice, Apprentice VS. Proficient, Distinguished Distribution of Biserial Correlations for Open-Response Items, Grade 4 87
Table 6-5: Novice, Apprentice, Proficient VS. Distinguished Distribution of Biserial Correlations for Open Response Items, Grade 4 88
Table 6-6: Distribution of Percent Correct For Multiple-Choice Items, Grade 4 89
Table 6-7: Distribution of Biserial Correlations for Multiple-Choice Items, Grade 4 90
Table 6-8: Distribution of Item Scores for Open-Response Items, Grade 8 91
Table 6-9: Distribution of Item-Total Correlations for Open-response Items, Grade 8 92

Table 6-10: Novice VS. Apprentice, Proficient, Distinguished Distribution of Biserial Correlations for Open-Response Items, Grade 8 93
Table 6-11: Novice, Apprentice VS. Proficient, Distinguished Distribution of Biserial Correlations for Open-Response Items, Grade 8. 94
Table 6-12: Novice, Apprentice, Proficient VS. Distinguished Distribution of Biserial Correlations for Open Response Items, Grade 8 95
Table 6-13: Distribution of Percent Correct for Multiple-Choice Items, Grade 8 96
Table 6-14: Distribution of Biserial Correlations for Multiple-Choice items, Grade 8 97
Table 6-15: Distribution of Item Scores for Open-Response Items, Grade 12 98
Table 6-16: Distribution of Item-Total Correlations for Open response Items, Grade 12 99
Table 6-17: Novice VS. Apprentice, Proficient, Distinguished Distribution of Biserial Correlations for Open Response Items, Grade 12 100
Table 6-18: Novice, Apprentice VS. Proficient, Distinguished Distribution of Biserial Correlations for Open Response Items, Grade 12 101
Table 6-19: Novice, Apprentice, Proficient VS. Distinguished Distribution of Biserial Correlations for Open Response Items, Grade 12 102
Table 6-20: Distribution of Percent Correct for Multiple-Choice Items, Grade 12 103
Table 6-21: Distribution of Biserial Correlations for Multiple-Choice Items, Grade 12 104
Table 7-1: Rules for Determining Students for whom a School is Accountable 106
Table 7-2: Grade 4 Students in Each Accountability Eligibility Category 108
Table 7-3: Grade 8 Students in Each Accountability Eligibility Category 109
Table 7-4: Grade 12 Students in Each Accountability Eligibility Category 110
Table 7-5: Testing Windows 111
Table 7-6: Profile of Performance Events Facilitator Qualifications and Demographics 114
Table 8-1: Number of Scorers and Training Leaders at each Grade 119
Table 8-2: Profile of Scorer Qualifications and Demographics 120
Table 8-3: Inter-Rater Reliability in Scoring of Open-Response Questions 122
Table 8-4: Student-Level Inter-Rater Correlations 122
Table 9-1: Number of 1993-94 Matrix and Common Items Administered as Matrix Items in 1992-93 128
Table 9-2: Number of Item Difficulties Identified as Outliers 128
Table 9-3: Item Difficulty Statistics and Adjustment Coefficients 130
Table 9-4: Impact of Adjusting Cut Scores Used To Assign Performance Levels on Matrix-Samples Items 132
Table 9-5: Crosstabulation of 1993-94 Rescoring of Grade 12 Reading Responses With Scores Originally Assigned in 1992-93, in Percent 133
Table 9-6: Rescoring Adjustments for 1993-94 Abilities 133
Table 9-7: Percentages of Students Classified as Distinguished on Open-Response Items 135
Table 9-8: Number of Tasks Completed by Grade, Content Area, and Year 137
Table 9-9: Distributions of Students Across Performance Level by Grade and Content Area 139

Table 9-10: Percentage, By Task, Of Students at Each Score Point for Grade 4 Mathematics Performance Events 140
Table 9-11: Percentage of Students at Each Score Point 140
Table 9-12: Performance Levels Obtained By Equating Sample Students for 1991-92 and 1993-94 Performance Events 142
Table 9-13: Percentage of Kentucky Students at Each Performance Level, 1993-94 Performance Events 159
Table 11-1: Statewide Accountability Index Scores 147
Table 11-2: Statewide Accountability Index Scores, By Region 148
Table 11-3: Accountability Index Scores By Content and Year 149
Table 12-1: Comparison of Original Scores to Rescores for Inconsistent Teachers Grade 4 156
Table 12-2: Comparison of Original Scores to Rescores for Inconsistent Teachers Grade 8 157
Table 12-3: Comparison of Original Scores to Rescores for Inconsistent Teachers Grade 12 157
Table 12-4: Qualifying Process Results 159
Table 12-5: Results of Second Qualifying Round 159
Table 12-6: Number of Portfolios Scored in Random Sample of Schools 160
Table 12-7: Comparison of Original Scores From Randomly Selected Schools to Rescores Grade 4 . 161
Table 12-8: Comparison of Original Scores From Randomly Selected Schools to Rescores Grade 8 . 161
Table 12-9: Comparison of Original Scores From Randomly Selected Schools To Rescore Grade 12 . 162
Table 12-10: Percentages of 1992-93 Writing Portfolios Rescored into Various Categories for Inconsistent Teachers and the Statewide Random Sample Grade 4 163
Table 12-11: Percentages of 1992-93 Writing Portfolios Rescored into Various Categories for Inconsistent Teachers and the Statewide Random Sample Grade 8 164
Table 12-12: Percentages of 1992-93 Writing Portfolios Rescored into Various Categories for Inconsistent Teachers and the Statewide Random Sample Grade 12 164
Table 12-13: Statewide Averages on Writing Portfolios 167
Table 12-14: Comparison of Original and Summer Rescoring Results for Audit Schools and Other Sampled Schools 170
Table 12-15: Consensus Portfolio Rescoring Activity 172
Table 12-16: Consensus Portfolio Results 173
Table 12-17: Comparison of Scoring Indices Based on Consensus Portfolios 174
Table 12-18: Writing Portfolio Audit Review Options 179
Table 12-19: Writing Portfolio Performance Level Assignments by Different Scoring Groups 180
Table 12-20: Levels of Agreement for Audit Review Scores with Original Scores and with Audit Scores 181
Table 12-21: Qualifying Rates 186
Table 12-22: Moderation Results 187
Table 12-23: Quality Control Portfolio Scoring Results 187
Table 12-24: Qualifying Results 189

Table 12-25: Portfolio Monitoring Results 189
Table 12-26: Quality Control Results 189
Table 12-27: Double-Score Consistency Results 190
Table 12-28: Scoring Accuracy and Student Performance 191
Table 12-29: Recast Writing Portfolio Results 193
Table 12-30: 1993-94 Portfolio Score Accuracy Statistics For Schools Audited in 1992-93 194
Table 13-1: Open-Response Test Reliabilities by Subject and Grade 197
Table 13-2: Preliminary Results: Generalizability Coefficients Treating Students and Tasks as Fixed or Random 198
Table 14-1: Grade 4 Student-Level Correlations Among Common Item Scores 202
Table 14-3: Grade 12 Student-Level Correlations Among Common Item Scores 203
Table 14-4: Grade 4 Student-Level Correlations, Lack of Good Faith Effort Scores Removed 206
Table 14-5: Grade 8 Student-Level Correlations, Lack of Good Faith Effort Scores Removed 207
Table 14-6: Grade 12 Student-Level Correlations, Lack of Good Faith Effort Scores Removed 207
Table 14-7: Grade 4 School-Level Correlations 208
Table 14-8: Grade 8 School-Level Correlations 208
Table 14-9: Grade 12 School-Level Correlations 209
Table 14-10: Correlation Between CTBS NCES and KIRIS Open-Response Questions, 1991-92 211
Table 14-11: Numbers of Schools and Students Obtained From NAEP Sampling, and Correlations Between NAEP and KIRIS Mathematics 212
Table 14-12: Correlation of KIRIS Open-Response Test Raw Scores With Act Scale Scores 214
Table 15-1: School-Level Correlations of Demographic Variables and Accountability Indices 225
Table 15-2: Reward/Assistance Classification By Grade Level 226
Table 15-3: Percent of Variance Explained In Percent of Improvement Goal, By Accountability Grade 227
Table 15-4: Correlation of Baseline and Percent of Improvement Goal, By Grades 228
Table 15-5: Correlation of Percent Free/Reduced Lunch and Percent of Improvement Goal, By Grade 228
Table 15-6: Rankings By District Achievement and District Wealth, (From the *Herald-Leader*) 230
Table 15-7: Correlation of Racial/Ethnic Minority Enrollment and Percent of Improvement Goal, By Grade 235
Table 16-1: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 4 Reading 237
Table 16-2: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 4 Social Studies 237

Table 16-3: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 4 Math 238
Table 16-4: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 4 Science 238
Table 16-5: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 8 Reading 239
Table 16-6: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 8 Math 239
Table 16-7: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity Grade 8 Social Studies 240
Table 16-8: Comparison of Mean Scores on Open-Response and Multiple-Choice Test By Gender and Ethnicity, Grade 8 Science 240
Table 16-9: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 12 Reading 241
Table 16-10: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 12 Social Studies 241
Table 16-11: Comparison of Mean Scores on Open-Response and Multiple Choice Tests By Gender and Ethnicity, Grade 12 Science 242
Table 16-12: Comparison of Mean Scores on Open-Response and Multiple-Choice Tests By Gender and Ethnicity, Grade 12 Math 242

TABLE OF FIGURES

Figure 9-1: Difficulties for 1993-94 Mathematics Items Re-Used From 1993-92, Grade 12 Plot of 1992-93 Difficulties (DIF93) With 1993-94 Difficulties (DIF94).. 125
Figure 9-2: Difficulties for Matrix-Sampled 1993-94 Mathematics Items Re-Used From 1993-92, Grade 12 Plot of 1992-93 Difficulties (DIF93) With 1993-94 Difficulties (DIF94).. 126
Figure 9-3: Difficulties for Matrix-Sampled 1993-94 Reading Items Re-Used From 1993-92, Grade 4 Plot of 1992-93 Difficulties (DIF93) With 1993-94 Difficulties (DIF94).. 129
Figure 12-1: Average 1992-93 Kentucky Writing Portfolio Indices Based on Scoring Standards as Implemented by Different Sets of Scorers.. 176
Figure 15-1: States With NSF-Sponsored Systemic Reform Initiatives in Mathematics and Science Whose Assessment Systems Meet Certain Criteria.. 221
Figure 15-2: Per-Pupil Revenue, Seek School Funding.. 234

TABLE OF APPENDICES

Appendix 7-1: Principal's and District Assessment Coordinator's Certification of Proper Test Administration Forms.. . . . . A-1
Appendix 7-2: Student Answer Document... . . . . A-2
Appendix 7-3: Summary of Performance Events. A-3
Appendix 9-1: Locations of Open-Response Linking Items by Year A-4
Appendix 9-2: Open-Response Test Raw Score to Performance Level Conversion Tables A-5
Appendix 9-3: Performance Event Raw Score to Performance Level Conversion Tables.. . . . . A-6
Appendix 10-1: September 1994 School Report... . . . . A-7
Appendix 10-2: January 1995 School Report... . . . . A-8

SECTION I
CONTEXT

Evaluation of any program must be in terms of its goals and purposes. To this end, the first section of this technical manual describes the purpose and contents of this manual (Chapter 1) and the background of the testing program and accountability system (Chapter 2).

CHAPTER 1
INTRODUCTION

The purpose of this technical manual is to provide in one place all major existing information about the technical characteristics of the Kentucky Instructional Results Information System (KIRIS), the assessment component of the Kentucky Education Reform Act (KERA) Accountability System between the 1991-92 and 1993-94 school years. While some parts of this manual may be used by educated laypersons, its intended audience is experts in psychometrics and educational research. It assumes working knowledge of measurement concepts such as reliability and validity, as well as basic familiarity with advanced topics in measurement and statistics such as graded-item-response models, generalizability theory, multi-trait, multi-method analysis, and factor analysis.

COMPREHENSIVENESS

Despite the several hundred pages of tables and prose in this manual, it is by no means intended to provide all available detail about the testing program or accountability system. To do so would have made the technical manual prohibitively expensive. Instead, in addition to summaries of important studies and references throughout the manual, we have included a bibliography of papers, publications, and books that may not have been included in the reference section.

YOU CAN'T COMPLETE THE EVALUATION OF AN EGG BEFORE THE CHICKEN HATCHES

Numerous hours of discussion were spent considering what should be included in this manual. The instructional consequences of the various choices made in developing and implementing the testing and accountability systems was our primary consideration (as long as the accuracy of the accountability index remained acceptably high). Yet, it is not possible to gather much information about instructional consequences in advance of implementing the accountability system, since it is the system that engenders the consequences. Moreover, most attempts at educational change require many years before their impacts are clearly visible. First, teachers must understand the system, then they must decide how to react to it. With feedback from the assessment and accountability programs based on a single testing cycle each year, it seems likely that it will take teachers several iterations before they can decide how best to meet the needs of their particular communities and students. Then, a sufficient number of students must experience the changed curriculum and instruction for enough time for it to show an effect. This is a necessary weakness of any technical documentation that can be produced now. Nonetheless, it is incumbent upon us to provide as much analysis as possible of instructional consequences at this time, even if such analyses are incomplete or are inconclusive. We have attempted, to the best of our abilities, to do so.

NOT JUST A TEST

In reading this technical manual, it is critical to remember that the testing program does not exist in a vacuum. It is one part of a complex network intended to help schools focus their energies on dramatic improvement in student learning. That is, KIRIS is part of an integrated program of testing, accountability, and curricular and instructional support. It can only be evaluated properly in this full context.

CHAPTER 2
BACKGROUND

To understand the Kentucky testing and accountability programs, one must understand the context from which they arose. This chapter describes the court case and ensuing executive and legislative actions that led to the current system.

KENTUCKY CONSTITUTION

The Kentucky Constitution, adopted in 1891, states that the "*General Assembly shall, by appropriate legislation, provide for an efficient system of common schools throughout the state.*"

COUNCIL FOR BETTER EDUCATION V. WILKINSON

In November 1985, the Council for Better Education, a nonprofit corporation formed by 66 school districts, seven boards of education, and 22 public school children sued the state for not providing an efficient system of education. In October 1988, Franklin County Circuit Court judge Roy Corns found for the plaintiffs. Seven capacities establishing a new legal framework for the school curriculum were set forth by the court as necessary for an adequate education:

> *Schools must provide (i) significant oral and written communication skills to enable students to function in a complex and rapidly changing civilization; (ii) sufficient knowledge of economic, social, and political systems to enable the student to make informed choices; (iii) sufficient understanding of government processes to enable the student to understand the issues that affect his or her community, state, and nation; (iv) sufficient self-knowledge and knowledge of his or her mental and physical wellness; (v) sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage; (vi) sufficient training or preparation for advanced training in either academic or vocational fields so as to enable each child to choose and pursue life work intelligently; and (vii) sufficient levels of academic or vocational skills to enable public school students to compete favorably with their counterparts in surrounding states, in academics or in the job market. Council for Better Education v. Wilkinson, NO. 85-CI-1759, slip op. at 4 (Franklin Cir. Ct., Oct. 14, 1988).*

In June 1989, the Kentucky Supreme Court directed the General Assembly to recreate and reestablish a new "efficient system of common schools" that complied with the Kentucky Constitution. The Court defined an efficient system of common schools as "an organization that provides a free and adequate education to all students throughout the state regardless of geographical location or local fiscal resources."

COUNCIL ON SCHOOL PERFORMANCE STANDARDS

In February, 1989, on his own initiative, Governor Wallace Wilkinson issued an executive order creating a twelve-member Council on School Performance Standards. Charged with determining what all students should know and be able to do and how learning should be assessed, the Council began its work by traveling across the Commonwealth conducting focus group interviews with business leaders, employers of graduates, parents, and educators, and asking them what a high school graduate in the year 2000 should know and be able to do. Information from the group interviews was used to construct a twenty-two minute telephone survey. More than 830 Kentucky residents of voting age were called and asked their opinions about what should be expected of future high school graduates.

The responses from the nearly 300 hours of telephone interviews were clear and consistent. While there was plenty of support for learning basic communications and mathematics skills, the most important result of learning identified was that students should be able to use their skills and knowledge in mathematics, science, social studies, and other disciplines in situations similar to what they will experience in the real world. In addition, Kentucky residents believed students should develop personal attributes such as self-sufficiency, self-confidence, and adaptability. They should learn to become responsible members of a family, work group, and community as part of their school learning experience. Also, Kentuckians believed students should learn to think and solve problems, make informed choices, and support their decisions. The information from the statewide survey was presented to five task forces of educators from across the state who were asked to advise the Council on possible learning goals for all students and ways in which these goals might be assessed.

In September, 1989, the Council on School Performance Standards produced the report *Preparing Kentucky Youth for the Next Century: What Students Should Know and Be Able To Do and How Learning Should Be Assessed* and presented it to the Curriculum Committee of the Legislative Task Force charged with creating Kentucky's new "efficient system of common schools." Six broad learning goals for all students were recommended with particular emphasis on what they should be able to do. These learning goals are presented in Table 2-1.

TABLE 2-1
KENTUCKY'S SIX LEARNER GOALS

1. Students are able to use basic communication and mathematics skills for purposes and situations they will encounter throughout their lives.
2. Students shall develop their abilities to apply core concepts and principles from mathematics, the sciences, the arts, the humanities, social studies, practical living studies, and vocational studies to what they will encounter throughout their lives.
3. Students shall develop their abilities to become self-sufficient individuals.
4. Students shall develop their abilities to become responsible members of a family, work group, or community, including demonstrating effectiveness in community service.
5. Students shall develop their abilities to think and solve problems in a variety of situations they will encounter in life.
6. Students shall develop their abilities to connect and integrate experiences and new knowledge from all subject matter fields with what they have previously learned and build on past learning experiences to acquire new information through media sources.

In addition, the Council recommended that the state launch a major effort to assess student performance beyond what can be measured by paper-and-pencil tests. It also was recommended that the state initiate long-range development efforts that support school reform in implementing the new learning goals.

KENTUCKY EDUCATION REFORM ACT

In 1990, the Council's recommendations were incorporated into House Bill 940, the Kentucky Education Reform Act (KERA), as a first step in redefining the school curriculum and providing what the courts required as an "adequate education for all students." The bill was signed by Governor Wallace Wilkinson on April 11, 1990 and became law on July 13, 1990.

With KERA, the General Assembly established the framework for a major revision of Kentucky's educational system. KERA required the establishment of learning goals for the educational system, provided a procedure by which those goals would be defined and assessed, and created a series of rewards and assistance to be associated with performance of schools on those assessments. Table 2-2 presents the text of one portion of the Kentucky Education Reform Act, KRS 158.645 – *Capacities Required of Students in Public Education*, which is a modification of language used in the court decision.

TABLE 2-2
KRS 158.645 – CAPACITIES REQUIRED OF STUDENTS IN PUBLIC EDUCATION.

The General Assembly recognizes that public education involves shared responsibilities. State government, local communities, parents, students, and school employees must work together to create an efficient school system. Parents and students must assist schools with efforts to assure student attendance, preparation for school, and involvement in learning. The cooperation of all involved is necessary to ensure that desired outcomes are achieved. It is the intent of the General Assembly to create a system of public education which shall allow and assist all students to acquire the following capacities:

- *(1) Communication skills necessary to function in a complex and changing civilization;*
- *(2) Knowledge to make economic, social, and political choices;*
- *(3) Understanding of governmental processes as they affect the community, the state, and the nation;*
- *(4) Sufficient self-knowledge and knowledge of his mental and physical wellness;*
- *(5) Sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage;*
- *(6) Sufficient preparation to choose and pursue his life's work intelligently; and*
- *(7) Skills to enable him to compete favorably with students in other states.*

Kentucky's new reform act, however, went beyond adopting six learning goals. Elements of the law that have a direct effect on the curriculum and what students learn include the following:

1. *Schools shall require students to demonstrate a high level of achievement on each of the six learning goals.*
2. *Schools shall increase their attendance rates, reduce drop-out and retention rates, reduce physical and mental barriers to learning, and increase students' successful transition to post-high school experiences.*
3. *The Council on School Performance Standards shall frame the six school goals in measurable terms that define expected student learning outcomes. Statewide committees of practitioners shall be involved in this task.*
4. *The State Board for Elementary and Secondary Education shall create and implement a statewide, primarily performance-based assessment program.*
5. *All schools in the Commonwealth shall be held accountable for the proportion of students who achieve the school goals at the expected level set by the State Board.*

These requirements of the Reform Act placed special conditions on how the Council was to frame the school goals in measurable terms and how it would redefine what students learn in school. First, since there would be high stakes accountability (known as rewards and sanctions for schools) and teachers would be "teaching toward the test," the test must be worth "teaching to." Therefore, the focus of learning must be on what students are able to do with their knowledge and skills. Second, since teachers are to be teaching toward goals which are framed in measurable terms, the goals must contain what is "essential" and most important for students to learn. Third, since the school is accountable for all students, the goals should describe what is expected for all students. It is important to note that the Council decided the academic expectations, which would define the goals in measurable terms, should not be defined as minimum competencies, but rather as standards of performance for educated citizens.

The Kentucky Educational Reform Act also called for the establishment of a committee of national testing experts to design the accountability system.

ACADEMIC EXPECTATIONS

To further define what was expected of students, in December 1991 the State Board adopted 75 statements, at that time called Valued Outcomes, describing what was expected of students. During the first three years of testing, more than 40 Valued Outcomes were incorporated into the new assessment system. In 1994 it became clear that the language used in the 75 Valued Outcomes did not clearly convey their intent to many important audiences, including parents, legislators, media commentators, and some educators. In addition, there were concerns both within and outside the Department of Education regarding the near-term feasibility of measuring Learner Goal 3 (self-sufficiency) and Learner Goal 4 (responsible group membership). To address both sets of concerns the Kentucky Department of Education (KDE) met with or received comments from more than 175 people including parents, educators, students, representatives of business and government, clergy, and non-affiliated citizens. Comments were received from both critics and advocates of Kentucky's education reform efforts.

The result of this process was the establishment of 57 academic expectations that describe what Kentucky students should be able to know and do when they graduate from high school. Table 2-3 presents the 12 academic expectations (the original outcomes 5-9 were combined into one academic expectation) related to Goal 1 – Communication Skills. Table 2-4 presents the 37 expectations (the original outcomes 16 and 21 were combined) related to Goal 2 – Core Academic Concepts. Table 2-5 presents the five expectations for Goal 5 –Think and Solve Problems, and Table 2-6 presents the three expectations for Goal 6 – Integrating Knowledge.

TABLE 2-3
ACADEMIC EXPECTATIONS FOR GOAL 1: COMMUNICATION SKILLS

1. Students use reference tools such as dictionaries, almanacs, encyclopedias, and computer reference programs, and reference tools such as interviews and surveys to find the information they need to meet specific demands, explore interests, or solve specific problems.

2. Students make sense of the variety of materials they read.

3. Students make sense of the variety of things they observe.

4. Students make sense of the various messages to which they listen.

5-9. Students use mathematical ideas and procedures to communicate, reason, and solve problems.

10. Students organize information through development and use of classification rules and systems.

11. Students write using appropriate forms, conventions, and styles to communicate ideas and information to different audiences for different purposes.

12. Students speak using appropriate forms, conventions, and styles to communicate ideas and information to different audiences for different purposes.

13. Students make sense of ideas and communicate ideas with the visual arts.

14. Students make sense of ideas and communicate ideas with music.

15. Students make sense of ideas and communicate ideas with movement.

16. Students use computers and other kinds of technology to collect, organize, and communicate information and ideas.

TABLE 2-4
ACADEMIC EXPECTATIONS FOR GOAL 2: CORE ACADEMIC CONCEPTS

SCIENCE

1. Students understand scientific ways of thinking and working and use those methods to solve real-life problems.
2. Students identify, analyze, and use patterns such as cycles and trends to understand past and present events and predict possible future events.
3. Students identify and analyze systems and the ways their components work together or affect each other.
4. Students use the concept of scale and scientific models to explain the organization and functioning of living and nonliving things and predict other characteristics that might be observed.
5. Students understand that under certain conditions nature tends to remain the same or move toward a balance.
6. Students understand how living and nonliving things change over time and the factors that influence the changes.

MATHEMATICS

7. Students understand number concepts and use numbers appropriately and accurately.
8. Students understand various mathematical procedures and use them appropriately and accurately.
9. Students understand space and dimensionality concepts and use them appropriately and accurately.
10. Students understand measurement concepts and use measurements appropriately and accurately.
11. Students understand mathematical change concepts and use them appropriately and accurately.
12. Students understand mathematical structure concepts including the properties and logic of various mathematical systems.
13. Students understand and appropriately use statistics and probability.

SOCIAL STUDIES

14. Students understand the democratic principles of justice, equality, responsibility, and freedom and apply them to real-life situations.
15. Students can accurately describe various forms of government and analyze issues that relate to the rights and responsibilities of citizens in a democracy.
16. Students observe, analyze, and interpret human behaviors, social groupings, and institutions to better understand people and the relationships among individuals and among groups.
17. Students interact effectively and work cooperatively with the many ethnic and cultural groups of our nation and world.
18. Students understand economic principles and are able to make economic decisions that have consequences in daily living.
19. Students recognize and understand the relationship between people and geography and apply their knowledge in real-life situations.
20. Students understand, analyze, and interpret historical events, conditions, trends, and issues to develop historical perspective.
21. (Incorporated into 2.16)

ARTS AND HUMANITIES

22. Students create works of art and make presentations to convey a point of view.
23. Students analyze their own and others' artistic products and performances using accepted standards.
24. Students have knowledge of major works of art, music, and literature and appreciate creativity and the contributions of the arts and humanities.
25. In the products they make and the performances they present, students show that they understand how time, place, and society influence the arts and humanities such as languages, literature, and history.
26. Though the arts and humanities, students recognize that although people are different, they share some common experiences and attitudes.
27. Students recognize and understand the similarities and differences among languages.
28. Students understand and communicate in a second language.

PRACTICAL LIVING

29. Students demonstrate skills that promote individual well-being and healthy family relationships.
30. Students evaluate consumer products and services and make effective consumer decisions.
31. Students demonstrate the knowledge and skills they need to remain physically healthy and to accept responsibility for their own physical well-being.
32. Students demonstrate strategies for becoming and remaining mentally and emotionally healthy.
33. Students demonstrate the skills to evaluate and use services and resources available in their community.
34. Students perform physical movement skills effectively in a variety of settings.
35. Students demonstrate knowledge and skills that promote physical activity and involvement in physical activity throughout their lives.

VOCATIONAL STUDIES

36. Students use strategies for choosing and preparing for a career.
37. Students demonstrate skills and work habits that lead to success in future schooling and work.
38. Students demonstrate skills such as interviewing, writing resumes, and completing applications that are needed to be accepted into college or other postsecondary training or to get a job.

TABLE 2-5
ACADEMIC EXPECTATIONS FOR GOAL 5: THINK AND SOLVE PROBLEMS

1. Students use critical thinking skills such as analyzing, prioritizing, categorizing, evaluating, and comparing to solve a variety of problems in real-life situations.
2. Students use creative thinking skills to develop or invent novel, constructive ideas or products.
3. Students organize information to develop or change their understanding of a concept.
4. Students use a decision-making process to make informed decisions among options.
5. Students use problem-solving processes to develop solutions to relatively complex problems.

TABLE 2-6
ACADEMIC EXPECTATIONS FOR GOAL 6: INTEGRATE KNOWLEDGE

1. Students connect knowledge and experiences from different subject areas.
2. Students use what they already know to acquire new knowledge, develop new skills, or interpret new experiences.
3. Students expand their understanding of existing knowledge by making connections with new knowledge, skills, and experiences.

SELECTING A CONTRACTOR

On April 19, 1991 the Kentucky Department of Education issued a request for proposals for a contractor to design and implement an assessment and accountability program consistent with the Kentucky Education Reform Act. Six bids were received and judged by the panel of nationally recognized experts that had been established under KERA to design the accountability system. Table 2-7 presents the names of the members of a national panel of experts who reviewed the proposal, and Table 2-8 presents the membership of the Kentucky Technical Review Committee. Table 2-9 presents the bidders and judgment each review panel made about each proposal. Based on this process, the national experts and the Kentucky Department of Education recommended to the Kentucky Board of Education that the contract be awarded to Advanced Systems in Measurement and Evaluation, Inc. (ASME).

TABLE 2-7
NATIONALLY RECOGNIZED EXPERTS
WHO ASSESSED THE PROPOSALS

Skip Kifer
Jason Millman
Doris Redfield
Doug Rindone
Grant Wiggins

TABLE 2-8
KENTUCKY TECHNICAL REVIEW COMMITTEE

Roger Pankratz
Frank Yeager
Booker Rice
Jack Foster
Ken Olsen
Ben Oldham
Wendy Birdwell
Steven Henderson
Diane Teasley
H. M. Snodgrass

TABLE 2-9
BIDDERS AND JUDGMENTS ABOUT THEIR PROPOSALS

Bidder	National Experts	Kentucky Technical Review Committee Rankings (1 is highest)	
		Technical Merit	Overall*
Advanced Systems in Measurement and Evaluation	Best	3	6
CTB MacMillan/McGraw-Hill	Not responsive	2	3
Educational Testing Service	Acceptable	1	4
Measurement Incorporated	Not responsive	4	2
Psychological Corporation	Not responsive	5	5
School Research and Service	Not responsive	6	1

*Weights used in determining the overall ranking strongly favored dollar amount of bid rather than technical merit.

After analyzing the proposals, review committee comments, and references from past clients, KDE recommended to the Board that Advanced Systems in Measurement and Evaluation be awarded the contract. On July 3, 1991, the State Board voted to accept the recommendation.

THE ACCOUNTABILITY SYSTEM

KIRIS data are used to hold schools accountable. This is done to heighten public attention in order to focus schools on helping their students achieve the high standards set by committees of Kentucky educators. Inherent in the accountability system is the recognition that the massive changes in instruction and learning will require many years to be achieved. Although there is no definite schedule, many have spoken of this as an effort that will take at least 20 years.

In order to keep schools focused throughout this lengthy period, a system of setting subgoals was developed. To reflect the distinctive starting points of different schools and their different populations, a unique improvement goal is calculated for each school to keep it moving in the right direction. Data from one testing period form the operational basis of implementing the policy decisions about setting each school's improvement goal.

KERA called for the integration of all information collected into three statistics for each school in the state. First, a ***baseline***, or initial accountability index, is computed. From this baseline, an ***improvement goal*** (in previous publications the improvement goal was referred to as the "threshold") is set. Each school's improvement goal represents a 10% improvement against the difference between its baseline and 100. Finally, a school's ***accountability index*** is calculated and compared to its improvement goal. For example, if a school's baseline is 30, the difference between its baseline and 100 is 70. Ten percent of 70 is 7. Thus, the school's improvement goal is 30+7, or 37.

Specifically, for the first accountability cycle, 1991-92 testing data provided the baseline against which the combined results of 1992-93 and 1993-94 testing are compared. Similarly, the 1992-93 and 1993-94 testing will provide a new baseline against which the 1994-95 and 1995-96 results will be compared.

NONCOGNITIVE DATA. In addition to the assessment data collected through **on-demand testing, performance events,** and **portfolios** (for descriptions of each testing method, see Chapter 3, Test Specifications), KERA called for the accountability system to include certain noncognitive indicators of school success:

- attendance (used in all grades)
- retention (used in all grades)
- drop-out rate (used in grades 8 and 12 only)
- transition to adult life (used in grade 12 only)
- reduction of physical and mental barriers to learning (currently being studied for future use)

As with the cognitive areas, an advisory committee was created to ensure the input of Kentucky educators in the design of this component of the accountability system. Using the criteria called for by KERA as its foundation, the committee guided the development of the definitions of each variable and made suggestions for alterations in the system for future years. For example, a sub-committee is currently exploring definitions and procedures for collecting data pertaining to the reduction of physical and mental health barriers to learning, one of the indicators stipulated in KERA.

SECTION 2

TEST DEVELOPMENT AND ITEM ANALYSIS

The quality of the inferences made from test scores depends on the quality of the test development and standard setting processes. Item analysis provides us with objective evidence regarding the success of those efforts. Chapter 3 presents the frameworks used to ensure that each year's set of test forms is parallel to those of other years. Chapter 4 describes the test development process used to ensure that all tests and items are of very high quality. Chapter 5 describes how standards were set to describe the four Kentucky performance categories. Chapter 6 summarizes the results of item analyses.

CHAPTER 3
TEST SPECIFICATIONS

There are two primary reasons why test specifications are necessary. First, they help ensure that the test is measuring what it is intended to measure. Second, test specifications help ensure that for each year within a biennium, the sample of tasks upon which school success is judged is equivalent.

Two sets of specifications have been used to develop KIRIS assessments. A subset of the 57 academic expectations that explicate four of Kentucky's six learner goals forms one set of specifications. Because the law requires KIRIS to be like the National Assessment of Educational Progress (NAEP)[1], a modification of the NAEP content frameworks forms the second set of specifications. Based on these two frameworks, Kentucky's assessment program targets large concepts and themes; the connecting of ideas; and the application of knowledge, skills, and reasoning to solve problems.

In addition to the aforementioned specifications, Kentucky's curriculum framework provides examples and further information about what kinds of questions might be asked in the KIRIS tests. To provide an additional level of detail, KDE disseminated draft detailed content guidelines for all KIRIS tests in fall 1994. While these guidelines were first used to build the 1994-95 assessments, retroactively the most recent tests in the first accountability cycle were checked against them for compatibility.

LEARNER GOALS

The genesis of Kentucky's six learner goals is presented in Chapter 2. Table 3-1 lists the four goals that are the basis of the assessment program. (In Table 2-1 these are referred to as Goals 1, 2, 5 and 6; this numbering scheme is followed throughout the manual to avoid confusion.) For the first accountability cycle, the tests concentrated on the first two learner goals.

TABLE 3-1
KENTUCKY'S FOUR LEARNER GOALS THAT ARE MEASURED IN KIRIS

1. Students are able to use basic communication and mathematics skills for purposes and situations they will encounter throughout their lives.
2. Students shall develop their abilities to apply core concepts and principles from mathematics, the sciences, the arts, the humanities, social studies, practical living studies, and vocational studies to what they will encounter throughout their lives.
5. Students shall develop their abilities to think and solve problems in a variety of situations they will encounter in life.
6. Students shall develop their abilities to connect and integrate experiences and new knowledge from all subject matter fields with what they have previously learned, and build on past learning experiences to acquire new information through media sources.

---

[1] Although the law requires the transitional KIRIS to be NAEP-like, it does not define the characteristics of NAEP that should be embodied in KIRIS. Since the law also requires KIRIS to be primarily performance-based, KDE has defined NAEP-like to mean having substantially similar content frameworks and to be statistically linked for the purpose of inferring national norms.

ACADEMIC EXPECTATIONS

The history of the 57 academic expectations is described in Chapter 2. Demonstrating that these expectations are met requires students to show that they can use skills and knowledge in situations they might experience in life (e.g., writing a business letter, producing a report, making an oral or video presentation. Solving a problem, creating a new product, or choosing and supporting the best course of action). That is, an academic expectation is a statement of student ability to demonstrate consistent quality performances on authentic tasks related to a skill area, core concept, personal attribute, or thinking process. Consistent quality performance means a high level of performance on a variety of tasks over time.

At this time, the statewide assessment does not reflect all fifty-seven academic expectations. Because the scope of reform was so comprehensive, it would have been unfair to hold all schools in Kentucky immediately responsible for the academic expectations which require special programs or facilities. Additionally, the Kentucky Board of Education determined that the academic expectations in Learner Goals 3 and 4 (self-sufficiency and responsible group membership) were too difficult to assess at this time. Therefore, the KIRIS assessment does not reflect any academic expectations from those goals. Likewise, the academic expectations found in goals 5 and 6 have not been explicitly targeted for assessment (though many of the performance and portfolio items require the skills identified under these goals).

Tables 3-2 through 3-6 show which academic expectations were targeted with the open-response tests in each grade and in each testing year. Tables 3-7 through 3-11 present similar information based on the performance events tests. Tables 3-12 through 3-14 note the distribution of open-response items across academic expectations. Because the open-response questions tend to be broad and often belong to multiple categories (i.e., address multiple academic expectations), totaling the numbers of items in the categories within a major subject area will usually yield a larger number than the actual count of items. This "double counting" of questions might lead to vastly disproportionate numbers of items in an area. This is especially true in science, where large numbers of items fall into the patterns and systems categories. (Some would agree that science is the study of patterns.) Data and procedures (operations) in mathematics also tend to "capture" many items. Some categories are considerably overlapping form one another by definition - e.g., measurement and data, democratic principals and political systems, social systems and cultural diversity. Some academic expectations may "define" a smaller portion of a content domain than others - e.g., mathematical structure. (Tables 3-11 through 3-16 are based on data provided by Stuart Kahl at Advanced Systems).

TABLE 3-2
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 1
OPEN RESPONSE TESTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
1.1 Accessing Information									
1.2 Reading	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.3 Observing									
1.4 Listening									
1.5-1.9 Math Procedures			✓			✓			✓
1.10 Classifying									
1.11 Writing									
1.12 Speaking									
1.13 Visual Arts									
1.14 Music									
1.15 Movement									
1.16 Technology									

TABLE 3-3
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 2, OPEN RESPONSE TESTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
2.1 Scientific Skills	✓	✓	✓	✓	✓	✓	✓	✓	✓
2.2 Patterns	✓	✓	✓		✓	✓	✓	✓	✓
2.3 Systems and Interactions	✓	✓	✓	✓	✓	✓	✓	✓	✓
2.4 Models and Scales	✓		✓	✓	✓	✓	✓	✓	✓
2.5 Constancy		✓	✓	✓	✓	✓		✓	✓
2.6 Evolutionary Change		✓	✓		✓	✓		✓	✓
2.7 Number Concepts	✓		✓		✓	✓		✓	✓
2.8 Mathematical Procedures	✓	✓	✓	✓	✓	✓		✓	✓
2.9 Space and Dimensionality	✓	✓	✓			✓	✓		✓
2.10 Measurement		✓		✓	✓	✓		✓	✓
2.11 Change		✓	✓	✓	✓	✓	✓	✓	✓
2.12 Mathematical Structure			✓		✓	✓	✓		✓
2.13 Data		✓	✓			✓		✓	✓
2.14 Democratic Principals		✓		✓				✓	✓
2.15 Political Systems	✓		✓		✓	✓			✓
2.16 Social Systems			✓	✓	✓	✓		✓	✓
2.17 Cultural Diversity			✓		✓	✓	✓		✓
2.18 Economic Systems		✓	✓		✓	✓			✓
2.19 Geography		✓	✓	✓		✓			✓
2.20 Historical Perspective	✓	✓	✓			✓	✓	✓	✓
2.22 Production									
2.23 Analysis of Form									
2.24 Aesthetics									
2.25 Cultural Heritage									
2.26 Cultural Diversity									

TABLE 3-3, continued
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 2, OPEN RESPONSE TESTS

Academic Expectation	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
2.27 Language									
2.28 Second Language									
2.29 Family Life		✓	✓					✓	
2.30 Consumerism		✓	✓		✓	✓		✓	✓
2.31 Physical Wellness		✓	✓		✓	✓		✓	✓
2.32 Mental Wellness									
2.33 Health Systems			✓						
2.34 Psychomotor Skills									
2.35 Physical Activities					✓	✓			
2.36 Career Path		✓	✓		✓	✓		✓	✓
2.37 Transition Skills		✓	✓		✓	✓			
2.38 Post Secondary Search					✓	✓		✓	✓

TABLE 3-4
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 5
OPEN RESPONSE TESTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
5.1 Critical Thinking									
5.2 Creative Thinking									
5.3 Conceptualizing									
5.4 Decision Making			✓						
5.5 Problem Solving			✓			✓			

TABLE 3-5
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 6
OPEN RESPONSE TESTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
6.1 Multiple Perspectives									
6.2 Acquire Knowledge									
6.3 Expanding Knowledge									

TABLE 3-6
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 1, PERFORMANCE EVENTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
1.1 Accessing Information									
1.2 Reading									
1.3 Observing									
1.4 Listening									
1.5-1.9 Math Procedures									
1.10 Classifying									
1.11 Writing									
1.12 Speaking									
1.13 Visual Arts			✓			✓			✓
1.14 Music						✓			✓
1.15 Movement			✓						
1.16 Technology									

TABLE 3-7
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 2, PERFORMANCE EVENTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
2.1 Scientific Skills	✔	✔	✔	✔	✔	✔	✔	✔	✔
2.2 Patterns	✔	✔	✔	✔	✔	✔	✔	✔	✔
2.3 Systems and Interactions	✔	✔	✔	✔	✔	✔	✔	✔	✔
2.4 Models and Scales	✔		✔	✔	✔	✔	✔	✔	✔
2.5 Constancy	✔			✔			✔	✔	✔
2.6 Evolutionary Change			✔			✔	✔	✔	✔
2.7 Number Concepts	✔			✔			✔		✔
2.8 Mathematical Procedures	✔	✔		✔	✔			✔	✔
2.9 Space and Dimensionality	✔	✔		✔	✔	✔	✔	✔	✔
2.10 Measurement	✔	✔	✔	✔		✔	✔	✔	
2.11 Change	✔	✔	✔	✔	✔		✔	✔	✔
2.12 Mathematical Structure			✔				✔		✔
2.13 Data	✔	✔	✔	✔	✔	✔	✔	✔	✔
2.14 Democratic Principals				✔		✔		✔	
2.15 Political Systems	✔	✔		✔	✔	✔		✔	✔
2.16 Social Systems	✔	✔	✔			✔	✔		✔
2.17 Cultural Diversity	✔		✔	✔	✔	✔	✔	✔	✔
2.18 Economic Systems	✔	✔		✔		✔	✔	✔	✔
2.19 Geography	✔	✔	✔	✔	✔	✔	✔	✔	✔
2.20 Historical Perspective	✔			✔	✔	✔	✔	✔	✔
2.22 Production						✔			✔
2.23 Analysis of Form									
2.24 Aesthetics									
2.25 Cultural Heritage									
2.26 Cultural Diversity			✔						

TABLE 3-8
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 2, PERFORMANCE EVENTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
2.27 Language									
2.28 Second Language									
2.29 Family Life									
2.30 Consumerism		✓			✓				✓
2.31 Physical Wellness			✓		✓	✓		✓	
2.32 Mental Wellness			✓			✓			
2.33 Health Systems								✓	✓
2.34 Psychomotor Skills									
2.35 Physical Activities		✓							
2.36 Career Path					✓	✓			✓
2.37 Transition Skills		✓	✓					✓	✓
2.38 Post Secondary Search					✓	✓		✓	

TABLE 3-9
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 5
PERFORMANCE EVENTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
5.1 Critical Thinking	✓	✓	✓	✓	✓	✓	✓	✓	✓
5.2 Creative Thinking	✓	✓	✓	✓	✓	✓	✓	✓	✓
5.3 Conceptualizing									
5.4 Decision Making	✓	✓	✓	✓	✓	✓	✓	✓	✓
5.5 Problem Solving	✓	✓	✓	✓	✓	✓	✓	✓	✓

TABLE 3-10
ASSESSMENT OF KENTUCKY'S ACADEMIC EXPECTATIONS - GOAL 6
PERFORMANCE EVENTS

ACADEMIC EXPECTATION	GRADE 4			GRADE 8			GRADE 12		
	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994	1991-1992	1992-1993	1993-1994
6.1 Multiple Perspectives	✓	✓	✓	✓	✓	✓	✓	✓	✓
6.2 Acquire Knowledge									
6.3 Expanding Knowledge									

TABLE 3-11
DISTRIBUTION OF ITEMS ACROSS ACADEMIC EXPECTATIONS
GRADE 4 OPEN-RESPONSE

Content Area	Academic Expectation	1991-92			1992-93			1993-94		
		Common	Matrix	Total	Common	Matrix	Total	Common	Matrix	Total
Reading	1.2 Literature	2	5	7	2	11	13	2	11	13
	1.2 Information	1	4	5	1	7	8	1	5	6
	1.2 Practical	0	3	3	1	3	4	1	3	4
	1.2 Personal				1	3	4	1	2	3
Science	2.1 Nat./Invest	2	4	6	1	4	5	2	4	6
	2.2 Patterns	1	5	6	1	15	16	3	11	14
	2.3 Systems	1	8	9	2	12	14	2	7	9
	2.4 Models/Scales	2	1	3	0	2	2	1	3	4
	2.5 Consistency	0	2	2	2	4	6	1	5	6
	2.6 Evolution	0	1	1	1	1	2	0	3	3
	Inquiry	2	3	5	1	6	7	2	7	9
	Life	0	4	4	1	17	18	2	11	13
	Earth/Space	1	2	3	1	2	3	1	6	7
	Physical	0	3	3	2	2	4	1	6	7
Mathematics	2.7 Number	1	1	2	0	3	3	1	2	3
	2.8 Procedures	1	2	3	0	5	5	0	5	5
	2.9 Space/Dim.	1	1	2	1	3	4	0	3	3
	2.10 Measure	0	2	2	1	4	5	1	3	4
	2.11 Change	0	2	2	1	5	6	2	4	6
	2.12 Structure	0	1	1	0	0	0	0	2	2
	2.13 Data	0	3	3	2	4	6	1	4	5
Social Studies	2.14 Democratic	0	1	1	1	0	1	0	0	0
	2.15 Political	0	0	0	0	6	6	2	7	9
	2.16 Social	1	0	1	0	3	3	0	4	4
	2.17 Cultural	0	1	1	0	1	1	0	1	1
	2.18 Economic	0	1	1	1	3	4	1	2	3
	2.19 Geology	0	3	3	1	2	3	1	2	3
	2.20 History	1	2	3	1	6	7	1	5	6
		3 Common, 12 Matrix-sampled per area			5 Common, 24 Matrix-sampled per area			5 Common, 24 Matrix-sampled per area		

TABLE 3-12
DISTRIBUTION OF ITEMS ACROSS ACADEMIC EXPECTATIONS
GRADE 8 OPEN-RESPONSE

Content Area	Academic Expectation	1991-92			1992-93			1993-94		
		Common	Matrix	Total	Common	Matrix	Total	Common	Matrix	Total
Reading	1.2 Literature	1	5	6	2	7	9	1	6	7
	1.2 Information	1	1	2	1	8	9	2	8	10
	1.2 Practical	0	3	3	2	5	7	1	3	4
	1.2 Personal	1	3	4	1	4	5	1	4	5
Science	2.1 Nat./Invest	3	5	8	2	4	6	2	5	7
	2.2 Patterns	2	9	11	3	10	13	2	10	12
	2.3 Systems	1	6	7	4	17	21	3	18	21
	2.4Models/Scales	3	3	6	2	10	12	2	8	10
	2.5 Consistency	0	3	3	2	11	13	2	9	11
	2.6 Evolution	0	5	5	1	10	11	2	10	12
	Inquiry	1	4	5	2	3	5	1	4	5
	Life	1	4	5	2	9	11	1	10	11
	Earth/Space	0	2	2	0	6	6	2	6	8
	Physical	1	2	3	1	6	7	1	3	4
Mathematics	2.7 Number	0	3	3	1	5	6	1	3	4
	2.8 Procedures	0	2	2	1	5	6	1	6	7
	2.9 Space/Dim.	1	1	2	0	4	4	1	3	4
	2.10 Measure	0	3	3	2	4	6	1	4	5
	2.11 Change	1	0	1	0	1	1	0	1	1
	2.12 Structure	1	0	1	1	3	4	0	4	4
	2.13 Data	0	3	3	0	5	5	1	5	6
Social Studies	2.14 Democratic	0	2	2	0	2	2	1	2	3
	2.15 Political	0	2	2	0	1	1	0	2	2
	2.16 Social	0	2	2	1	3	4	1	3	4
	2.17 Cultural	1	2	3	0	3	3	0	4	4
	2.18 Economic	0	3	3	1	3	4	1	3	4
	2.19 Geology	1	3	4	0	4	4	1	4	5
	2.20 History	2	2	4	0	5	5	1	5	6
		3 Common, 12 Matrix-sampled per area			5 Common, 24 Matrix-sampled per area			5 Common, 24 Matrix-sampled per area		

TABLE 3-13
DISTRIBUTION OF ITEMS ACROSS ACADEMIC EXPECTATIONS
GRADE 12 OPEN-RESPONSE

Content Area	Academic Expectation	1991-92			1992-93			1993-94		
		Common	Matrix	Total	Common	Matrix	Total	Common	Matrix	Total
Reading	1.2 Literature	1	5	6	1	8	9	2	7	9
	1.2 Information	1	1	2	2	6	8	1	7	8
	1.2 Practical	0	3	3	2	4	6	1	4	5
	1.2 Personal	1	3	4	2	6	8	1	4	5
Science	2.1 Nat./Invest	3	5	8	2	14	16	3	14	17
	2.2 Patterns	2	9	11	5	18	23	3	19	22
	2.3 Systems	1	6	7	3	15	18	3	14	17
	2.4Models/Scales	3	3	6	1	11	12	2	12	14
	2.5 Consistency	0	3	3	1	11	12	3	10	13
	2.6 Evolution	0	5	5	3	9	12	2	9	11
	Inquiry	1	4	5	2	7	9	2	7	9
	Life	1	4	5	2	6	8	1	5	6
	Earth/Space	0	2	2	1	6	7	1	6	7
	Physical	1	2	3	0	5	5	1	5	6
Mathematics	2.7 Number	0	3	3	1	2	3	1	2	3
	2.8 Procedures	0	2	2	1	6	7	0	5	5
	2.9 Space/Dim.	1	1	2	0	5	5	1	5	6
	2.10 Measure	0	3	3	2	3	5	1	3	4
	2.11 Change	1	0	1	1	3	4	1	3	4
	2.12 Structure	1	0	1	0	3	3	1	4	5
	2.13 Data	0	3	3	1	7	8	0	4	4
Social Studies	2.14 Democratic	0	2	2	1	3	4	1	3	4
	2.15 Political	0	2	2	0	2	2	0	3	3
	2.16 Social	0	2	2	1	3	4	1	3	4
	2.17 Cultural	1	2	3	0	6	6	1	4	5
	2.18 Economic	0	3	3	0	2	2	1	2	3
	2.19 Geology	1	3	4	0	3	3	1	2	3
	2.20 History	2	2	4	2	3	5	0	4	4
		3 Common, 12 Matrix-sampled per area			5 Common, 24 Matrix-sampled per area			5 Common, 24 Matrix-sampled per area		

TABLE 3-14
DISTRIBUTION OF ITEMS ACROSS ACADEMIC EXPECTATIONS
GRADE 4 MULTIPLE-CHOICE

Content Area	Academic Expectation	1991-92			1992-93			1993-94		
		Common	Matrix	Total	Common	Matrix	Total	Common	Matrix	Total
Reading	1.2 Literature	23	73	96	12	50	62	4	44	48
	1.2 Information	11	45	56	6	26	32	6	27	33
	1.2 Practical	7	24	31	4	11	15	5	11	16
	1.2 Personal	4	14	18	3	9	12	4	8	12
Science	2.1 Nat./Invest	6	27	33	6	23	29	5	23	28
	2.2 Patterns	23	123	146	14	48	62	8	39	47
	2.3 Systems	14	81	95	6	37	43	7	36	43
	2.4Models/Scales	9	22	31	4	11	15	1	10	11
	2.5 Consistency	16	62	78	4	31	35	4	25	29
	2.6 Evolution	1	8	9	0	3	3	0	8	8
	Inquiry	9	41	50	6	24	30	5	22	27
	Life	12	49	61	7	28	35	6	27	33
	Earth/Space	10	46	56	5	21	26	5	22	27
	Physical	9	44	53	5	28	33	5	29	34
Mathematics	2.7 Number	8	38	46	3	16	19	3	15	18
	2.8 Procedures	9	36	45	4	13	17	3	12	15
	2.9 Space/Dim.	6	25	31	3	12	15	3	11	14
	2.10 Measure	6	39	45	3	19	22	3	19	22
	2.11 Change	1	0	1	2	7	9	3	8	11
	2.12 Structure	4	24	28	2	9	11	2	9	11
	2.13 Data	6	30	36	3	20	23	3	20	23
Social Studies	2.14 Democratic	4	11	15	2	9	11	2	9	11
	2.15 Political	5	29	34	3	15	18	2	14	16
	2.16 Social	4	16	20	2	7	9	2	9	11
	2.17 Cultural	5	12	17	2	8	10	2	9	11
	2.18 Economic	5	27	32	3	13	16	4	14	18
	2.19 Geology	6	30	36	3	15	18	3	12	15
	2.20 History	5	26	31	3	14	17	2	15	17

TABLE 3-15
DISTRIBUTION OF ITEMS ACROSS ACADEMIC EXPECTATIONS
GRADE 8 MULTIPLE-CHOICE

Content Area	Academic Expectation	1991-92			1992-93			1993-94		
		Common	Matrix	Total	Common	Matrix	Total	Common	Matrix	Total
Reading	1.2 Literature	13	64	77	11	33	44	9	34	43
	1.2 Information	15	44	59	5	33	38	8	35	43
	1.2 Practical	10	24	34	9	17	26	4	14	18
	1.2 Personal	7	24	31	44	13	17	4	13	17
Science	2.1 Nat./Invest	9	27	36	2	19	21	6	19	25
	2.2 Patterns	16	90	106	12	40	52	11	31	42
	2.3 Systems	16	86	102	6	40	46	8	36	44
	2.4Models/Scales	17	33	50	6	18	24	2	22	24
	2.5 Consistency	7	53	60	7	21	28	2	21	23
	2.6 Evolution	6	20	26	3	11	14	2	10	12
	Inquiry	10	36	46	5	24	29	6	20	26
	Life	11	46	57	5	24	29	4	23	27
	Earth/Space	9	45	54	5	24	29	5	24	29
	Physical	10	53	63	5	24	29	5	29	34
Mathematics	2.7 Number	7	19	26	3	12	15	3	11	14
	2.8 Procedures	11	44	55	3	16	19	5	15	20
	2.9 Space/Dim.	4	21	25	3	16	19	2	16	18
	2.10 Measure	6	42	48	4	16	20	3	15	18
	2.11 Change	4	13	17	3	15	18	3	17	20
	2.12 Structure	4	9	13	1	5	6	1	7	8
	2.13 Data	4	32	36	3	16	19	3	15	18
Social Studies	2.14 Democratic	3	15	18	2	9	11	2	12	14
	2.15 Political	4	16	20	2	9	11	2	11	13
	2.16 Social	5	11	16	4	7	11	3	7	10
	2.17 Cultural	3	12	15	1	9	10	2	11	13
	2.18 Economic	3	21	24	3	13	16	2	13	15
	2.19 Geology	9	28	37	2	16	18	2	14	16
	2.20 History	4	26	30	2	14	16	4	14	18

TABLE 3-16
DISTRIBUTION OF ITEMS ACROSS ACADEMIC EXPECTATIONS
GRADE 12 MULTIPLE-CHOICE

Content Area	Academic Expectation	1991-92			1992-93			1993-94		
		Common	Matrix	Total	Common	Matrix	Total	Common	Matrix	Total
Reading	1.2 Literature	15	63	78	8	33	41	7	27	34
	1.2 Information	14	43	57	6	27	33	8	33	41
	1.2 Practical	7	41	48	7	14	21	5	17	22
	1.2 Personal	10	33	43	4	22	26	5	19	24
Science	2.1 Nat./Invest	15	50	65	6	26	32	7	32	39
	2.2 Patterns	23	98	121	10	56	66	9	63	72
	2.3 Systems	16	70	86	7	33	40	8	44	52
	2.4Models/Scales	15	71	86	6	33	40	9	34	43
	2.5 Consistency	12	53	65	6	26	32	6	33	39
	2.6 Evolution	4	30	34	4	21	25	3	22	25
	inquiry	11	50	61	6	24	30	5	25	30
	Life	12	52	64	5	25	30	5	23	28
	Earth/Space	9	44	53	5	23	28	5	24	29
	Physical	11	46	57	4	24	28	5	24	29
Mathematics	2.7 Number	6	12	18	2	7	9	2	6	8
	2.8 Procedures	8	49	57	3	22	25	5	20	25
	2.9 Space/Dim.	6	29	35	3	12	15	3	14	17
	2.10 Measure	5	31	36	3	14	17	2	14	16
	2.11 Change	5	19	24	4	16	20	2	17	19
	2.12 Structure	5	12	17	2	9	11	3	10	13
	2.13 Data	5	28	33	3	18	21	3	17	20
Social Studies	2.14 Democratic	4	21	25	2	11	13	3	11	14
	2.15 Political	5	23	28	2	12	14	2	12	14
	2.16 Social	4	19	23	2	11	13	4	11	15
	2.17 Cultural	4	16	20	2	11	13	2	11	13
	2.18 Economic	4	26	30	2	11	13	3	12	15
	2.19 Geology	6	24	30	4	12	16	2	12	14
	2.20 History	4	26	30	3	14	17	2	14	16

CURRICULUM FRAMEWORK

The academic expectations describe what students are expected to do with the knowledge and skills they acquire in school. They do not define in detail all the knowledge and individual skills students need. The Kentucky Education Reform Act charged the Department of Education to develop a curriculum framework to be used by teachers and school administrators in designing local curriculum.

More than 100 teachers, counselors, administrators, regional service center consultants, and university personnel were significantly involved in the development of *Transformations: Kentucky's Curriculum Framework*. Volume I of the curriculum framework was designed to help administrators and teachers address the academic expectations in their local or school curricula. Each academic expectation is defined by a set of demonstrators which are benchmark or indicators of student progress toward the expectation. The demonstrators are accompanied by learning links and related concepts which provide teachers with ideas about how to incorporate the academic expectations in their curricula. Also included are sample activities, which suggest ways to involve students in engaging instructional/assessment experiences. The activities are coordinated with the demonstrators and the statewide assessment.

The publication of *Transformations* has provided Kentucky's educators with a document that integrates the processes of curriculum, instruction, and assessment. It was an important step in helping support the power of the assessment to affect instruction positively.

KIRIS NAEP-BASED FRAMEWORKS

The content domains for the major subject areas tested are described by the assessment frameworks, which detail lists of subdomains and cognitive processes within each subject area. The frameworks, based on those employed in creating NAEP tests, serve three purposes. First, they reflect an underlying rationale for the testing approach, based on current theory or research. Second, they guide the process of test development, helping to ensure that coverage of a subject area domain is balanced and complete. Third, they provide a definition of reporting categories — subdomains for which test results are produced and reported. In the following sections, the frameworks used for multiple-choice and open-response questions at accountability grades and for the voluntary continuous assessment tests (in the non-accountability grades) are presented in detail for reading, mathematics, science, social studies, writing, the arts and humanities, practical living, and vocational studies.

READING. The reading assessment framework is based on the principle that reading is an interaction among the reader, the text, and the situation. The reader interacts with the text, using prior knowledge and a repertoire of reading strategies gained through experience to make sense of and extend meaning from the text for a purpose. Kentucky's reading framework, similar to that of the 1992 NAEP framework, is based on this definition of reading and on Kentucky's academic expectation that all students will exhibit basic communication skills. The categories are described in Table 3-18. For each testing year in the accountability cycle, the set of test forms covers each indicated area in the table.

TABLE 3-17
KIRIS READING CONTENT FRAMEWORK

Item Category	Passage Type			
	Literary	Informa-tional	Practical/ Workplace	Persuasive
Comprehension: Making Sense of What was Read	✓	✓	✓	✓
Comprehension: Elaborating and Responding Critically	✓	✓	✓	✓
Reader Awareness	✓	✓	✓	✓
Reader Response	✓	✓	✓	✓

PASSAGE TYPE. The passages used in the assessment are drawn from authentic texts, rather than those written solely for the purpose of testing. The passages appeared with pictures and headings as originally published, when appropriate or available. The quality of the writing was another selection criterion.

Effective readers approach different kinds of materials in different ways based on the purposes in reading the type of text thus, passages were chosen from:

- literary pieces, such as excerpts from novels, short stories, poems, plays, letters or essays;
- informational pieces, such as articles from newspapers or magazines, content area textbook sections, or encyclopedias;
- practical/workplace pieces, such as transportation schedules, directions, or applications; and
- persuasive pieces, such as speeches, editorials, and advertisements.

Literary, informational, and practical passages are often associated with different purposes — reading for literary experience, reading to be informed, or reading to perform a task. The fourth passage type (persuasive) includes literary, informational, and practical passages whose purpose is largely to influence the opinion of readers.

Table 3-18 shows the approximate percentages of passage types used at each level for the open-response and multiple-choice tests that were administered each year during the first accountability cycle.

TABLE 3-18
PASSAGE TYPES USED IN THE KIRIS READING ASSESSMENT

Grade	Passage Type			
	Literary	Informational	Practical/ Workplace	Persuasive
4	50%	25%	15%	10%
8	35%	30%	20%	15%
12	30%	30%	20%	20%

ITEM CATEGORY. Effective readers demonstrate the ability to understand information presented in the text and to analyze and extend information based on their own experiences and prior knowledge. The framework for the Kentucky assessment reflects the belief that these behaviors should be practiced by readers at every level of development and should not be viewed as developing sequentially.

Making Sense of What was Read requires that readers formulate an initial, general impression of the material and continue to interpret the text as they read. The reader focuses on important portions of the text and links information from different sections by examination and analysis. The following questions and statements are typical of this process of making sense of what was read:

- What is the story about?
- Describe the main character.
- How did the character change from beginning to end?
- What is the article about?
- What caused this particular event?
- What is the author's position?
- What support does the author present for his/her position?

Elaborating and responding critically demands that readers connect their own knowledge with text and evaluate the text. Readers also extend, examine, analyze, apply, and judge the ideas in the text. The following questions demonstrate this category:

- How does this story compare/contrast with your own experiences?
- Would you recommend this story? Why or why not?
- Does this remind you of a current event? Explain.
- Compare this story/article to that one.
- What additional information would you need to complete this task?
- What would happen if you omitted step X?
- Do you agree with the author's position? Why or why not?

Reader Awareness. Effective readers demonstrate the ability to monitor and regulate their comprehension of text. Questions in this category addressed awareness/comprehension of reading strategies, such as:

- setting purposes for reading;
- previewing the text;
- identifying the type (genre) of the text;
- activating prior knowledge about the topic;
- visualizing while reading;
- using text structure and cues (e.g., italics);
- making predictions;
- paraphrasing and summarizing;
- seeking clarification;
- raising questions while reading;
- reading ahead or rereading;
- taking notes;
- reviewing comprehension.

Reader awareness questions were asked only in the multiple-choice portion of the tests and did not count towards school accountability scores.

Reader Attitude. Effective readers value reading and are likely to read frequently, which supports continued improvement of their reading skills. This component of the assessment examined students' perceptions about reading and about themselves as readers. Reader attitude questions were asked only in the multiple-choice portion of the tests and did not count towards school accountability scores.

MATHEMATICS. Table 3-19 shows the content framework used to build the mathematics tests. Because of the interconnected nature of mathematics (and the richness of each KIRIS open-response item), a single item often addresses more than one content and process category. In each year of the first accountability cycle, the set of mathematics tests at each grade level covers all content and process categories, with the exception of conceptual knowledge of numbers, which was not targeted for grade 4 in the 1992-93 test.

The content categories are the academic expectations for mathematics listed under "Goal 2 — Core Concepts and Principles" by Kentucky's Council on School Performance Standards. The process categories are from the *Curriculum and Evaluation Standards for School Mathematics* developed by the National Council of Teachers of Mathematics (NCTM). The entire framework is consistent with the most recent NAEP framework for mathematics.

The process categories, consistent with those described in the NCTM *Standards*, represent different thought processes or ways of mentally processing information. They are not mutually exclusive.

Test questions assessing students' conceptual knowledge in mathematics provide evidence of their abilities to label, verbalize, and define the concept; recognize and generate examples and nonexamples; use models, diagrams, and symbols to represent the concept; translate from one mode of representation to another; and compare, contrast, and integrate related concepts.

TABLE 3-19
KIRIS MATHEMATICS CONTENT FRAMEWORK

Content	Process			
	Conceptual Knowledge	Procedural Knowledge	Problem Solving	Reasoning & Analysis
Number	✓[2]	✓	✓	✓
Mathematical Procedures	✓	✓	✓	✓
Space & Dimensionality	✓	✓	✓	✓
Measurement	✓	✓	✓	✓
Change	✓	✓	✓	✓
Mathematical Structure	✓	✓	✓	✓
Data	✓	✓	✓	✓

Procedural knowledge requires students to demonstrate to reliably and efficiently execute procedures and verify their results empirically (e.g., using models) or analytically; verbalize and give reasons for the steps in a procedure; recognize correct and incorrect procedures; generate new procedures and extend or modify familiar ones; and recognize situations in which a procedure is appropriate, necessary, and correctly applied. Procedures include, but are not limited to, computational algorithms.

Students' ability to use their knowledge of mathematics for problem solving is measured by questions that required that they recognize and formulate problems by identifying the key features and questions in a given situation; solve problems by selecting and using appropriate strategies, models, and relevant mathematical content; and verify and generalize solutions by reflecting critically on the problem.

The assessment of mathematical reasoning and analysis requires students to use standard forms of reasoning such as inductive, deductive, and statistical inference; use contextual reasoning such as proportional and spatial reasoning; and identify and examine components of a given situation to abstract commonalities, specify properties and relationships, and identify underlying structure.

To clarify the mathematics content categories, Table 3-20 presents an overview of mathematics topics that belong to each one.

SOCIAL STUDIES. The development and selection of test questions for the social studies component of the assessment were guided by the two-dimensional matrix shown in Table 3-21.

---

[2] Conceptual knowledge of numbers was not targeted for grade 4 in the 1992-93 test.

TABLE 3-20
TOPICS ADDRESSED IN EACH MATHEMATICS CONTENT CATEGORY

Number	Mathematical Procedures	Space and Dimensionality
Number Relations Number Sense Infinity Notation Systems Estimation Number Theory Equivalent Forms Percents	Arithmetic Proportion Computational Algorithms Algebraic Processes	Regular and Irregular Figures in Various Dimensions Spatial Visualization Congruence Symmetry Transformations Similarity Coordinate Systems Vectors Orientation
Measurement	**Change**	**Mathematical Structure**
Direct and Indirect Measurements Maximum, Minimum Metric, Customary, and Nonstandard Units Length, Area, Volume, Capacity Mass, Weight Time, Rate Temperature Trigonometry Angle Curves Money	Transformations Algebraic, Matrix, and Trigonometric Representations Limit Series Continuous vs. Discrete Sequences Curve Fitting Functions Patterns Variables	Proof Fields Axiomatic Systems Logic Number Systems Order Relations Groups Discrete Structures Order of Operations
Data		
Applied Probability Theoretical Probability Probability in Experiments Measures of Central Tendency (including mean, median, mode) Measures of Dispersion (including range and outliers) Statistical Procedures and Decisions		

TABLE 3-21
KIRIS SOCIAL STUDIES CONTENT FRAMEWORK

Content	Cognitive Skills	
	Knowledge and Comprehension	Application and Higher Order
Democratic Principles	✓	✓
Structure and Function of Political Systems	✓	✓
Structure and Function of Social Systems	✓	✓
Cultural Diversity	✓	✓
Structure and Function of Economic Systems	✓	✓
Relationship of Geography to Human Activity	✓	✓
Historical Perspective	✓	✓
Intra- and Interpersonal Relationships		

The major content categories represent subdisciplines of social studies. The eight content categories listed are the social studies academic expectation categories identified by Kentucky's Council on School Performance Standards. The cognitive skills represent types of mental or cognitive processing consistent with Bloom's Taxonomy of Cognitive Objectives. These skills refer to one dimension of a question's characteristics, and the content categories refer to another. For example, a question requiring "higher order thinking skills" as well as an understanding of content would be placed in the appropriate content category (Democratic Principles, Cultural Diversity, etc.) and be identified as "higher order" according to the cognitive skill dimension. Social studies achievement results were grouped into broader categories for reporting.

The cognitive skills represent a hierarchy of mental processes required of students in answering the test questions. "Knowledge" questions require primarily memory processes, addressing lower-level skills such as factual recall. "Comprehension" questions assess the understanding of concepts and generally require students to demonstrate understanding by translating or explaining (or, in the case of multiple-choice questions, identifying appropriate translations or explanations). "Application" questions assess a slightly higher level of understanding by requiring students to apply their understanding of concepts to particular situations. Bloom's "higher order" processes include analysis, synthesis, and evaluation. Analysis questions require that information be broken down into parts so that ideas and their interrelationships are made explicit. Synthesis questions require that ideas be pulled together to produce a broader message or entity. Evaluation questions require students to make judgments about the relative value of ideas and approaches. In multiple-choice questions assessing this process, several response options often have some merit, but one is clearly "best" for any particular question.

SCIENCE. The conceptual framework used to guide the development of the science component of the assessment is shown in Table 3-22.

TABLE 3-22
KIRIS SCIENCE CONTENT FRAMEWORK

Science Content	Themes of Science					
	Nature of Scientific Activity	**Constancy**	**Evolution**	**Models and Scales**	**Patterns**	**Systems and Interactions**
Scientific Inquiry Nature of Science/Investigations Data Gathering/Reduction Data Analysis/Interpretation	✓	✓	✓	✓	✓	✓
Life Science Characteristics of Life Organisms Ecology/Environment	✓	✓	✓	✓	✓	✓
Earth/Space Science Astronomy Meteorology Geology/Natural Resources	✓	✓	✓	✓	✓	✓
Physical Science Matter Energy Force/Motion	✓	✓	✓	✓	✓	✓

The two major crossing dimensions, *Science Content* and *Themes of Science* include components of the 1994 NAEP science framework. *Scientific Inquiry* was added to the science content areas to accommodate questions on scientific inquiry which require little or very low levels of specific science content knowledge. This allowed for the reporting of results to reveal program strengths or weaknesses with respect to inquiry skills. Other questions involving inquiry skills, but which required a greater degree of content understanding as well, were included in their appropriate fields of science.

The themes of science are the science academic expectation categories of the Council on School Performance Standards. These themes cross the disciplines of science and are intended to help unify instruction and thinking across those fields. Because these themes are far from mutually exclusive (that is, a single question was likely to address several themes simultaneously), they are treated in the Kentucky framework just as they are by NAEP. Questions are counted under more than one theme for the reporting of results on the Council's academic expectations.

None of the dimensions so far described cleanly separate the questions by cognitive process levels. Therefore, questions were also identified as either *Knowledge/Comprehension* or *Application/Higher Order*. This assured that the questions in a particular category of another dimension were not overly dominated by either lower or higher order skills.

The science reporting categories define important concepts and skills within the different content categories of the framework.

WRITING. Proficiency in writing is measured using portfolios and an on-demand writing exercise (although for the first two accountability cycles, only the Writing Portfolio counts toward accountability decisions).

PORTFOLIOS. For portfolios, students select their best examples of writing within the content requirements set by the Kentucky Writing Advisory Committee. Those content requirements ensure that portfolios allow students to demonstrate that they meet the requirement of the pertinent academic expectation, *Students demonstrate the ability to communicate through their writing ideas and information to a variety of audiences for a variety of purposes in a variety of written forms.* Table 3-23 presents the entries required for each accountability grade level.

TABLE 3-23
KIRIS WRITING CONTENT REQUIREMENTS

Quality Criteria (All Grades)	Portfolio Pieces		
	Grade 4	Grade 8	Grade 12
Correctness (spelling, punctuation, capitalization) **Language** (usage, word choice) **Sentences** (correctness, structure, effectiveness) **Organization** (logic, coherence, transitions) **Idea Development and Support** **Purpose and Audience**	1. Table of Contents 2. A Letter to the Reviewer 3. One personal narrative 4. One short story, poem, play/script, or piece of fiction 5. One piece of writing the purpose of which is to: a. present/support a position, idea, or opinion; or b. tell about a problem and its solution; or c. inform 6. One piece of writing from a study area other than English/ Language Arts. 7. A Personal Selection	1. Table of Contents 2. A Letter to the Reviewer 3. One personal narrative 4-5. Two pieces of writing, each of which will achieve any one or more of the following purposes: a. predict an outcome b. defend a position c. analyze or evaluate a situation, person, place or thing d. solve a problem e. explain a process or concept f. draw a conclusion g. create a model 6. One short story, poem, play/script, or other piece of original fiction 7. A Personal Selection A minimum of one piece of writing must come from a subject area other than English or Language Arts.	1. Table of Contents 2. A Letter to the Reviewer 3. One personal narrative, memoir, personal essay, or personal vignette 4. One short story, poem, or play/script 5-7. Three pieces of writing, each of which will achieve any one or more of the following purposes: a. predict an outcome b. defend a position c. analyze or evaluate a situation, person, place or thing d. solve a problem e. explain a process or concept f. draw a conclusion g. create a model A minimum of two pieces of writing must come from a subject area other than English or Language Arts.

ON-DEMAND. A total of eighteen on-demand writing prompts were administered in 1991-92, with each student in grades 4, 8, and 12 responding to a prompt of his or her choosing from among the two prompts presented on each of nine forms. Three additional pairs of interdisciplinary writing prompts were developed for each grade level in 1992-93, with a total of twenty-four prompts being presented on twelve forms. Student responses were scored using the same scoring guide developed for writing portfolios, as described in Chapter 4.

The total set of prompts at each grade level covers all of the types of writing described in the portfolio section. Students choose to respond to one of the two prompts. Using a choice from two prompts makes the on-demand writing situation more closely resemble an authentic writing process. General directions encourage prewriting or planning and provide editing and revising suggestions that parallel the criteria by which the writing is evaluated. Prompts are chosen to be consistent with the amount of time students are able to sustain their writing: ninety minutes.

ARTS AND HUMANITIES. The framework for the assessment of arts and humanities is based on Kentucky's core concepts and related academic expectations. Table 3-24 presents the content framework.

TABLE 3-24
KIRIS ARTS AND HUMANITIES CONTENT FRAMEWORK

Academic Expectations	Area				
	Visual Arts	Music	Dance	Theater	Humanities
Create products and make presentations that convey concepts and feelings	✓	✓	✓	✓	
Analyze artistic products and performances	✓	✓	✓	✓	✓
Appreciate creativity and the values of the arts and the humanities	✓	✓	✓	✓	✓
Show an understanding of the influence of time, place, personality, and society on the arts and humanities	✓	✓	✓	✓	✓
Recognize differences and commonalities in the human experience	✓	✓	✓	✓	✓
Demonstrate awareness of the diversity of forms, structures, and concepts across languages and how they interrelate.	Not Assessed				
Understand and communicate in a second language.	Not Assessed				

The arts and humanities framework for all grade levels focuses on four major areas: the visual arts, music, dance, and drama[3]. Within those areas, the Arts and Humanities Advisory Committee identified specific knowledge and skills to be assessed. Table 3-25 presents additional information about the content of the Arts and Humanities test.

---

[3] Starting with the 1994-95 examination there will be an added emphasis in the humanities, especially literature.

TABLE 3-25
ADDITIONAL INFORMATION ABOUT THE CONTENT OF THE ARTS AND HUMANITIES TEST

Visual Arts	Music	Dance	Drama
1. Manipulation of medium 2. Six basic ways of making art: Drawing Painting Print Ceramics Sculpture Textiles 3. Elements, principles, design 4. Drawing from real life 5. Analysis 6. Appreciation 7. How do visual arts communicate meaning?	1. Basic understanding of elements: Rhythm Melody Harmony Form Expression of Style Texture Tonality 2. Know basic notations 3. Critique a work of past or present based on basic understanding 4. How is meaning communicated through music?	1. Elements of dance: Movement Space Time Force Technique Expression 2. Analysis 3. Appreciation – ability to communicate your own value system 4. How does dance communicate aesthetic value? (How is meaning portrayed through movement?)	1. Express through movement and expressions 2. Analyze a theater production (interpret) 3. How to role-play 4. Basic understanding of elements of creative theater

PRACTICAL LIVING AND VOCATIONAL STUDIES. The assessment frameworks developed for practical living and vocational studies are somewhat different from those used in the traditional academic areas but similar to the framework developed for the arts and humanities.

PRACTICAL LIVING. The practical living assessment addresses the development of health and fitness habits and behaviors; child and family development; consumer and resource management; and other concepts that support and will effect student growth, development, and wellness throughout their lives. It encompasses the traditional areas of health education, home economics, and physical education, but promotes both instruction and assessment that integrate those disciplines throughout the school years.

The practical living assessment is developed using core concepts based directly on pertinent academic expectations. These core concepts are presented in Table 3-26.

VOCATIONAL STUDIES. Vocational studies emphasizes the skills that all students need to make the successful transition from school to the world of work and to sustain careers that will enrich their lives.

The focus of the vocational studies assessment is further defined by school level. At the elementary level, students are evaluated on their developing appreciation of work, their awareness of themselves as workers, and their developing knowledge of career options. In the middle grades, the assessment focuses on students' investigations of career options, their knowledge of the relationships between careers and life roles, and their abilities to connect educational achievement to career opportunities. By the time they complete high school, students are expected to know the demands of a career and how it will affect their multiple roles in life. Additionally, students are to have acquired the knowledge and skills necessary for making a successful transition to the path they have selected for the immediate future – college, technical school, military service, or other work.

The core concepts and academic expectations addressed in the vocational studies assessment are presented in Table 3-26.

TABLE 3-26
KIRIS PRACTICAL LIVING/VOCATIONAL STUDIES CONTENT FRAMEWORK

Practical Living	Vocational Studies
Individual Well-Being and Healthy Family Relationships Consumer Skills Physical Wellness Mental and Emotional Wellness Community Service/Resource Systems Physical Movement Skills Lifetime Physical Activities	Choosing and Preparing for a Career Skills and Work Habits for Success in Future Schooling and Work Skills needed to be Accepted into College, Post-secondary Training, or job

ALTERNATE PORTFOLIOS. Students who are severely handicapped so that they are not expected to obtain a high school diploma are eligible to develop an Alternate Portfolio rather than participate in other aspects of the KIRIS program. Such students represent fewer than ½ of one percent of Kentucky's school population. Specific criteria for eligible students is detailed in KDE Program Advisory 92-OCAA-104 (February, 1993) and published in the *Kentucky Alternate Portfolio Teacher's Guide* and the annual *District Assessment Coordinator Implementation Guidebook.*

The Alternate Portfolio targets 28 Academic Expectations selected by the Alternate Portfolio Committee as being most assessable with a portfolio. A complete alternate portfolio includes:

- Table of Contents completed according to provided guidelines;
- Letter to the reviewer, describing the portfolio and its contents, written or dictated by the student (or a collaboration between the student and a non-disabled peer);
- Seven to ten entries representing the breadth of types, contexts, and domain areas;
  - One entry must be the student's weekly schedule and a description of its use, indicating the types of activities included, and the opportunities afforded for choice and for interaction with peers without disabilities;
  - At eighth and twelfth grades, one entry must be a résumé of volunteer and paid job experiences, accompanied by employer evaluations;
  - Photographs, audio- or videotapes, computer disks, and the like must be accompanied by a paragraph describing the entry and the rationale for its inclusion;
  - Student program data must be graphed and accompanied by a copy of the instructional program and a qualitative response from the student, non-disabled peer, and/or family member or caregiver;
  - One entry may be a sample of the student's present mode(s) of communication and description of its use;
- Letter from a family member or caregiver validating the contents of the portfolio.

ENTRY TYPES. Alternate portfolios include writing, projects, instructional program data, applications, and investigations or discoveries. Individual entries may span two or more types.

CONTEXT. Alternative Portfolio entries should demonstrate integration, functionality, age-appropriateness. They may use computer, assistive/adaptive, and other technology, and provide opportunities for choice.

DOMAIN. Appropriate domain areas include vocational, recreation or leisure, and personal management activities.

GROUP ENTRIES. The student may produce entries alone or as part of a group. Group entries must include an explanation and a reflection by the individual student.

CONTENT GUIDELINES

As the results of the first KIRIS assessment were being analyzed, it became apparent that educators across the Commonwealth needed additional information regarding the focus of the test items. This led to the development of detailed content guidelines which will be used both to guide the development of future test forms and to provide teachers with guidance about targeting their instruction. Copies of these draft Content Guidelines are available from KDE. The revised Content Guidelines will be released in August 1995.

ITEM TYPES

KERA mandated that the assessment program be "primarily performance-based." In 1990, no state had such an assessment program; therefore, it was not possible to implement the program by duplicating the work of others. The results of efforts conducted on smaller scales were used to guide the development of the program, but all such work needed refinement before it could be implemented on a statewide scale. While the program had to begin in the 1991-92 school year, it was clear that at least a five-year phase-in period was necessary to implement the assessment in its final form.

PORTFOLIOS. Kentucky writing and mathematics portfolios are compilations of five to seven best pieces of student work per portfolio. For 1993-94, all grade 4, 8, and 12 students enrolled in a Kentucky public school were required to complete both writing and mathematics portfolios (with minor exceptions as stipulated by the Kentucky Department of Education). Pieces for inclusion in these portfolios were produced throughout the school year as a part of the students' classroom curriculum. Schools' accountability indices for 1993-94 include writing portfolio scores, but mathematics portfolio scores are not included because they were not implemented in the baseline year.

The mathematics portfolio requires five to seven best pieces that demonstrate an understanding of the core concepts through the use of a variety of types of work and mathematical tools. The writing portfolio has specific requirements for each grade level that include a variety of types of work and content area requirements. Both the mathematics and writing portfolios require a Table of Contents and a Letter to the Reviewer reflecting on the student's work.

In the fall of each year, students were provided with Student Brochures that contained information about requirements, standards, and development issues. Teachers received teacher handbooks and training to assist them in working with students to develop portfolios.

When portfolios are completed, they are scored by Kentucky classroom teachers. To prepare for scoring, teachers receive the Holistic Scoring Guides, benchmark, and training portfolios for mathematics and/or writing. In addition, teachers received training in standards and scoring procedures.

In 1992-93, an Alternate Portfolio was instituted for students with severe handicapping conditions likely to prevent them from ever obtaining a high school diploma. The Alternate Portfolio is not part of the Accountability System for the first accountability cycle.

PERFORMANCE EVENTS. A typical performance event included a period of group work on a problem in a particular subject, followed by individual written work on the same problem. In all three years, performance events in mathematics, science, and social studies were administered for grades 4, 8, and 12. In 1992-93 and 1993-94 arts & humanities, and practical living/vocational studies performance events were also administered. In the latter two years, three tasks at each grade level were interdisciplinary; for example, requiring knowledge in and counting toward scores in both mathematics and science. The events were administered by performance events facilitators who brought materials to the schools, set them up, administered the events, and returned the response forms to Advanced Systems for scoring. In all but a few very small schools, each student took just one event in a group with a few other students. Table 3-27 presents an example of a performance event item. This Item, from Grade 12 Science, is typically presented as a sealed booklet. Instructions have been simplified to allow for presentation in this Technical Manual.

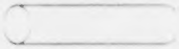
TABLE 3-27
EXAMPLE OF A PERFORMANCE EVENT ITEM

DOWNHILL RACERS

GENERAL INSTRUCTIONS:

Some objects will roll down a hill faster than other objects. You have been provided with a ramp and some materials to test. You will be working with one other student for the first 30 minutes of this task. After you have had 30 minutes to do your experiment, the test administrator will ask you to move apart and to finish the task on your own.

MATERIALS:
- 1 ramp
- 1 stop watch
- 1 of each of the items shown in the chart below

Objects in these 2 columns have the same diameter		Objects in these 2 columns have the same diameter	
 Glass Sphere	 Steel Sphere	 Glass Sphere	 Steel Sphere
 Solid Cylinder	 Solid Cylinder	 Solid Cylinder	 Solid Cylinder
 Hollow Cylinder	 Hollow Cylinder	 Hollow Cylinder	 Hollow Cylinder
Objects in this column all have the same mass	Objects in this column all have the same mass	Objects in this column all have the same mass	Objects in this column all have the same mass

BEST COPY AVAILABLE

TABLE 3-27, continued

1. You are to work as a pair and experiment to determine the time it takes for each of the materials to roll down the ramp. You will then determine the characteristics of the objects that make them roll faster. You will have 30 minutes to do this portion of your task. Both of you must keep a record of all the data from your experiment in the space provided below. When 30 minutes have gone by, the test administrator will have you separate and finish questions 2 through 4 by yourself.

DO NOT GO ON TO QUESTIONS 2 THROUGH 4 UNTIL YOU ARE TOLD TO DO SO.

You are to complete the following questions **by yourself when your test administrator tells you to do so.**

2. Based on the data you collected during step 1 of your investigation, what kind of objects roll fastest? Write a brief description of the variables you tested and the results of your investigations.

3. Design a new object that you think will roll faster than any of the objects you tested. Describe the size, shape and mass of the new object and explain why you think it will roll faster.

4. The test administrator will show you a new object. Raise your hand when you are ready to see the new object. Examine the new object and predict whether it will roll faster or slower than the objects you tested. Explain why you think so.

OPEN-RESPONSE QUESTIONS. Open-response questions were administered in reading, mathematics, science and social studies. In 1992-93 and 1993-94, some of the open-response items were considered interdisciplinary, incorporating concepts from the arts and humanities, practical living, or vocational studies. These interdisciplinary items received scores for each content area covered. Each open-response question is written so that students should be able to complete their responses in about ten minutes. Table 3-28 presents an example of an open-response item with a scoring guide. Table 3-29 gives an annotated student response to that same open-response question.

TABLE 3-28
EXAMPLE OF AN OPEN-RESPONSE ITEM, WITH SCORING GUIDE

1. Over a one-year time period, an insect population is known to quadruple. The starting population is fifteen insects.

   a. Make a table or graph to show the growth of the population from 0 through 6 year.

   b. How many insects would there be at the end of 10 years?

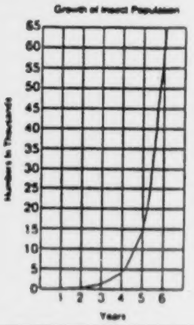
   c. Write a mathematical statement that would describe this growth.

   d. Would your statement or formula correctly describe the insect population after 50 years? Justify your conclusion.

Scoring Criteria for a 4-point response (on scale of 0-4)

The student gives correct answers to all questions, including a table or graph showing the strategy for finding the insect population, and complete explanation or justification for a change in population.

a. and b. Calculations	year 1	$15 \times 4^{1} =$	60
	year 2	$15 \times 4^{2} =$	240
	year 3	$15 \times 4^{3} =$	960
	year 4	$15 \times 4^{4} =$	3,840
	year 5	$15 \times 4^{5} =$	15,360
	year 6	$15 \times 4^{6} =$	61,440
	year 7	$15 \times 4^{7} =$	245,760
	year 8	$15 \times 4^{8} =$	983,040
	year 9	$15 \times 4^{9} =$	3,932,160
	year 10	$15 \times 4^{10} =$	15,728,640



c. at the end of n years. 15×4^{n}

d. valid reasons for some change in population

Open-response questions were included because they produce reliable scores, and they are reflective of KERA's educational direction. First, they require students to generate answers rather than merely recall or recognize information. Second, they emphasize student communication skills, and thus encourage teachers to use more powerful instructional practices that utilize communication skills as well as other aspects of problem solving.

TABLE 3-29
ANNOTATED RESPONSE TO OPEN-RESPONSE QUESTION

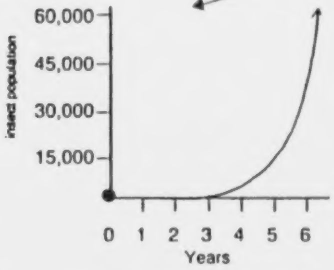
The student **accurately multiplies**
15x4,
15x4x4,
15x4x4x4,
15x4x4x4x4,
15x4x4x4x4x4,
and
15x4x4x4x4x4x4.

Quadruple is a mathematical term that the student recognizes and uses to **identify the rate of change** as 4. He or she might **relate to quadrant or quadrilateral** to identify that 4 is the "**multiplier.**"

The student extends the population by **multiplying** the sixth year, 61,440, by 4x4x4x4 or by **developing and evaluating an equation.**

a)

Years	Insects
0	15
1	60
2	240
3	960
4	3840
5	15,360
6	61,440



b) x = # of yrs of population growth
$15\,(4^x)$ = number of insects
$15\,(4^{10}) = 15{,}728{,}640$ insects

c) x = # of years
y = final # of insects
a = # of starting insects
$a\,(4^x) = y$

d) Depending on the insect, the formula would tell an approximate number of births of insects. However, it is unlikely that any insect can live past 50 and probably several hundred thousand of these will have died by then.

The student **organizes the data** in a chart or **shows graphing skills** to communicate the growth

The student **justifies the conclusion,** showing the application of real-world knowledge related to life-span, predators, food sources, etc., to weigh the **appropriateness of the mathematical projection** versus realistic populations over 50 years.

The student **develops a mathematical statement,** in words or as an algebraic formula, to **communicate the mathematical strategy** used.

MULTIPLE-CHOICE QUESTIONS. Multiple-choice questions have been written to require students to solve problems where the student recognizes the correct response. Each KIRIS multiple-choice question required about one minute for a typical student to respond. Scores from multiple-choice questions were used primarily for Chapter 1 evaluation reports, and were not used in computing a school's accountability index. Students also responded to a writing prompt, which typically requires about 90 minutes for a student to write a complete response.

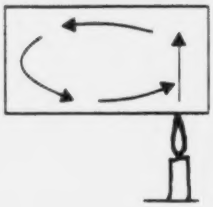
Despite their efficiency and ability to measure indirectly the ability of students to solve problems, multiple-choice questions do not closely model powerful instructional practices. In September 1994, the Kentucky Board of Education voted to eliminate multiple-choice questions from the KIRIS assessments starting with the 1994-95 testing year through at least 1996-97. During this time, the Department of Education will consider a number of ways to broaden the kinds of evidence of student performance collected as part of the testing program. Table 3-30 presents an example of a KIRIS multiple-choice question.

SUMMARY OF ITEM TYPES ADMINISTERED EACH YEAR. Changes over time include the emphases given to the different types of testing, the way testing is conducted, and the content areas to be included in each year. Tables 3-31, 3-32, and 3-33 provide a general description of the tests administered at grades 4, 8, and 12 in 1991-92, 1992-93, and 1993-94. Table 3-35 presents additional detail about interdisciplinary items.

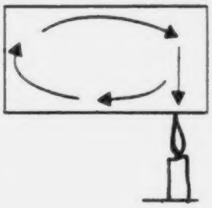
TABLE 3-30
EXAMPLE OF MULTIPLE-CHOICE ITEM

Which diagram shows the way water would move in a tank if you heated it as shown?

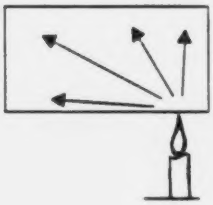
A.



B.



C.



D.

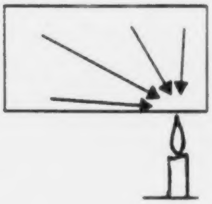


TABLE 3-31
TESTS ADMINISTERED IN GRADE 4 IN 1991-92, 1992-93, AND 1993-94
(CELL ENTRIES INDICATE TOTAL NUMBER OF UNIQUE ITEMS[4] ADMINISTERED ACROSS FORMS)

Subject Area	Portfolio			Performance Events			Open-Response[5]			Multiple-Choice		
	91-92	92-93	93-94	91-92	92-93	93-94	91-92	92-93	93-94	91-92	92-93	93-94
Mathematics		●	●	4	3[6]	3[6]	15	29	29	220	116	116
Reading				0	0	0	15	29	29	228	121	121
Science				4	4[6]	4[6]	15	29	29	220	116	116
Social Studies				4	4[6]	4[6]	15	29	29	220	116	116
Writing	●	●	●	0	0	0	18	24	24	0	0	0
Arts & Humanities				0	2	2	0	9	9	0	0	0
Practical Living / Vocational Studies				0	3	3	0	11	11	0	0	0
Alternate Portfolio		●	●	0	0	0	0	0	0	0	0	0
Total Unique Items				12	13	13	78	140	140	888	469	469

---

[4] Between 1991-92 and 1993-94, arts & humanities and practical living/vocational studies open-response scores were based on interdisciplinary items; such items are in the table two times.

[5] Writing open-response questions are different than those for other subject areas. See text for description.

[6] Includes interdisciplinary performance events (mathematics-science, mathematics-social studies, or science-social studies); such items are in the table two times.

TABLE 3-32
TESTS ADMINISTERED IN GRADE 8 IN 1991-92, 1992-93, AND 1993-94
(CELL ENTRIES INDICATE TOTAL NUMBER OF UNIQUE ITEMS[7] ADMINISTERED ACROSS FORMS)

Subject Area	Portfolio			Performance Events			Open-Response[8]			Multiple-Choice		
	91-92	92-93	93-94	91-92	92-93	93-94	91-92	92-93	93-94	91-92	92-93	93-94
Mathematics		●	●	4	3[9]	2[9]	15	29	29	220	116	116
Reading				0	0	0	15	29	29	228	121	121
Science				4	4[9]	3[9]	15	29	29	220	116	116
Social Studies				4	4[9]	4[9]	15	29	28	220	116	116
Writing	●	●	●	0	0	0	18	24	24	0	0	0
Arts & Humanities				0	3	3	0	6	9	0	0	0
Practical Living / Vocational Studies				0	4	4	0	11	11	0	0	0
Alternate Portfolio		●	●	0	0	0	0	0	0	0	0	0
Total Unique Items				12	15	14	78	140	139	888	469	469

---

[7] Between 1991-92 and 1993-94, arts & humanities and practical living/vocational studies open-response scores were based on interdisciplinary items; such items are in the table two times.

[8] Writing open response questions are different than those for other subject areas. See text for description.

[9] Includes interdisciplinary performance events (mathematics-science, mathematics-social studies, or science-social studies); such items are in the table two times.

TABLE 3-33
TESTS ADMINISTERED IN GRADE 12 IN 1991-92, 1992-93, AND 1993-94
(CELL ENTRIES INDICATE TOTAL NUMBER OF UNIQUE ITEMS[10] ADMINISTERED ACROSS FORMS)

Subject Area	Portfolio			Performance Events			Open-Response[11]			Multiple-Choice		
	91-92	92-93	93-94	91-92	92-93	93-94	91-92	92-93	93-94	91-92	92-93	93-94
Mathematics		●	●	4	4[12]	4[12]	15	29	29	220	116	116
Reading				0	0	0	15	29	29	228	121	121
Science				4	4[12]	4[12]	15	29	29	220	116	116
Social Studies				4	4[12]	4[12]	15	29	29	220	116	116
Writing	●	●	●	0	0	0	18	24	24	0	0	0
Arts & Humanities				0	2	3	0	9	9	0	0	0
Practical Living / Vocational Studies				0	3	4	0	11	11	0	0	0
Alternate Portfolio		●	●	0	0	0	0	0	0	0	0	0
Total Unique Items				12	16	17	78	140	140	888	469	469

---

[10] Between 1991-92 and 1993-94, arts & humanities and practical living/vocational studies open-response scores were based on interdisciplinary items; such items are in the table two times.

[11] Writing open-response questions are different than those for other subject areas. See text for description.

[12] Includes interdisciplinary performance events (mathematics-science, mathematics-social studies, or science-social studies); such items are in the table two times.

TABLE 3-34 INTERDISCIPLINARY OPEN-RESPONSE ITEMS ADMINISTERED IN 1992-93 AND 1993-94				
Additional Subject	Core Subject			
	Reading	Mathematics	Science	Social Studies
Arts and Humanities	3	1	2	3
Practical Living/Vocational Studies	2	4	3	2

NUMBER OF TEST QUESTIONS PER STUDENT

The technique of matrix sampling was used to increase the coverage of the assessment and to provide data for linking information across years. Matrix sampling divides a large number of test questions among different forms of the test; each student responds to only one form including all of the common items[13] and a fraction of the matrix-sampled items.

1991-92. In 1991-92, twelve forms of the test were created. For each subject area, each set of 12 forms contained a 180 multiple-choice questions and 12 open-response questions in addition to the common items. For mathematics, science, and social studies, the tests contained 40 common multiple-choice items and three common open-response items. The numbers of common multiple-choice reading questions were somewhat different. In addition to 40 reading comprehension questions, eight additional common multiple-choice questions focused on reading strategies[14]. The total number of reading questions was the same for all three grade levels assessed. The common items were publicly released following the administration of the tests, while the matrix-sampled items were kept secure. A subset of matrix-sampled items are reused the following year as common items. After being reused, these items become public with the rest of the common item pool.

1992-93 AND 1993-94. In 1992-93 and 1993-94, twelve forms of each test were created. For each subject area, each set of 12 forms covering a total of 96 multiple-choice questions and 24 open-response questions in addition to the common items. For mathematics, science, and social studies, the tests contained 20 common multiple-choice items and five common open-response items. The numbers of common multiple-choice reading questions were somewhat different. In addition to 20 reading comprehension questions, five additional common multiple-choice questions focused on reading strategies. The total number of reading questions was the same for all three grade levels assessed. The common items were publicly released following the administration of the tests, while the matrix-sampled items were kept secure. A subset of matrix-sampled items are reused the following year as common items. After being reused, these items become public with the rest of the common item pool.

Table 3-35 presents the number of common and matrix items within each subject taken by each student.

---

[13] Common items are those that are repeated in all test forms, and thus are taken by all students.

[14] Reading strategies questions did not count toward school accountability scores.

TABLE 3-35
NUMBERS OF KIRIS OPEN-RESPONSE AND MULTIPLE-CHOICE QUESTIONS FOR EACH SUBJECT PER STUDENT BY YEAR

Year	Question Type	Common Items		Matrix-Sampled Items (all subjects)
		Math, Science, and Social Studies	Reading	Per Student
1991-1992	Multiple-choice	40	48	15
	Open-response	3	3	1
1992-1993	Multiple-choice	20	25	8
	Open-response	5	5	2
1993-1994	Multiple-choice	20	25	8
	Open-response	5	5	2

CHAPTER 4
TEST DEVELOPMENT PROCESS

CONTENT ADVISORY COMMITTEES

The linchpins of the test development process are the content advisory committees that assure the tests and scoring guides are appropriate for Kentucky students. Advisory committees in each content area — reading, mathematics, science, social studies, writing, arts and humanities, and practical living/vocational studies — guide test development for open-response, multiple-choice, performance events, and the writing, mathematics, and alternate portfolios. Members of the committees include Kentucky classroom teachers at each pertinent grade level, resource teachers, statewide consultants, administrators, curriculum coordinators, university professors, and Kentucky Education Association representatives. Tables 4-1 and 4-2 summarize the number of committee members in 1993-94 and their backgrounds.

TABLE 4-1
TEST DEVELOPMENT COMMITTEES - MEMBER EXPERIENCE

Content Area	Teachers	Administrators	KDE Staff	Others
Mathematics	20	3	4	2
Reading	19	1	2	7
Science	18	0	4	3
Social Studies	20	0	5	2
Writing	20	7	12	2
Arts and Humanities	15	1	2	5
Practical Living/ Vocational Studies	19	1	5	3

THE TEST DEVELOPMENT SEQUENCE

Test development is an iterative process involving repeated reviews, revisions, and analyses. Throughout this process there is considerable interaction among content advisory committees, consultants from the Kentucky Department of Education, and staff from Advanced Systems in Measurement and Evaluation (or their subcontractors). As ASME and KDE gained experience, additional steps were added to the test development process in order to maximize test quality. Table 4-3 presents the major steps in the current test development process.

TABLE 4-2
TEST DEVELOPMENT COMMITTEES - MEMBER DEMOGRAPHICS[1]

Content Area	Male	Female	Minority	Non-Minority	Region							
					1	2	3	4	5	6	7	8
Mathematics	7	16	2	21	0	4	6	3	6	3	2	0
Reading	3	24	3	24	3	4	2	4	4	4	2	1
Science	7	18	3	22	3	2	4	2	3	1	2	1
Social Studies	9	17	3	23	4	7	3	2	1	3	1	1
Writing	6	35	3	38	3	4	4	5	12	5	4	4
Arts and Humanities	8	12	2	18	2	1	6	1	6	3	0	0
Practical Living/ Vocational Studies	12	11	4	19	4	3	1	5	5	3	1	1

1. ANALYZE SCORE DATA FROM PREVIOUS YEAR'S TEST. The last step of one year's test development process is the same as the first step of the next year's process. Item analysis provides information about how well each item performed. This information will be used to select old items that will appear in the next year's test. Table 4-4 describes the primary information produced for open-response items, and Table 4-5 shows a sample of the item analysis output for open-response items. Similar information is produced for multiple-choice items, excepting the biserial correlations as there are no performance level classifications associated with multiple choice items. Summaries of item analyses are presented in Chapter 6.

2. SELECT NEXT YEAR'S COMMON ITEMS FROM PREVIOUS YEAR'S MATRIX ITEMS. At their first meeting (usually in June), the content advisory committees select next year's common items. Decisions are based on both breadth of coverage of content and academic expectations and statistical quality of the items.

When committee members review test questions, they sort the items by content categories defined by the content guidelines and academic expectations. Advanced Systems' test developers and the Kentucky advisory committees hold joint responsibility to assure that the subset of questions within a category provide breadth of coverage of the important concepts and skills in that subdomain. With materials organized this way and with approximate quotas for the numbers of items in different categories required in a final test, the developers and committees easily maintain a relatively stable distribution and balance of item numbers across the categories.

3. SELECT EQUATING ITEMS FROM PREVIOUS YEAR'S MATRIX ITEMS. Similarly, at their first meeting the content advisory committees select matrix items that will be re-used to provide the equating link that maintains standards across test forms and year.

---

[1] Due to lost records, different sets of mutually exclusive categories may not add up to the total.

TABLE 4-3
MAJOR STEPS IN THE TEST DEVELOPMENT PROCESS

1.	Analyze score data from previous year's test
2.	Select next year's common items from previous year's matrix items
3.	Select equating items from previous year's matrix items
4.	Develop concepts for new items
5.	Write initial draft of new items
6.	Review and revise initial item drafts
7.	Review and select items for field testing (Content Advisory Committees)
8.	Assemble field test booklets
9.	Review field test booklets
10.	Administer field tests
11.	Analyze field test results
12.	Select new items for upcoming year's test forms
13.	Review selected items (Bias Review Committee)
14.	Review selected items (ASME staff)
15.	Review selected items (KDE staff)
16.	Revise selected items and assemble test forms
17.	Review each assembled form for content and sensitivity balance (ASME staff)
18.	Review each assembled form for content and sensitivity balance (KDE staff)
19.	Review final typeset copy for potential typographical errors

TABLE 4-4
ANALYSES PERFORMED ON OPEN-RESPONSE ITEMS

Analysis	Comments
1. Common item difficulty	Distribution of scores and average item score for the total group of examinees (on 0-4 scale).
2. Form item difficulty	Distribution of scores and average item score for common and matrix items based on group of examinees taking a particular form of the test (on 0-4 scale).
3. Item-total correlation	Correlation between item scores and scores on set of common items. For common items, this correlation is corrected for part-total contamination.
4. D_{N-A} biserial correlation	Discrimination between Novice and Apprentice. Biserial correlation between item score and student performance level, with performance level dichotomized as either Novice or above Novice.
5. D_{A-P} biserial correlation	Discrimination between Apprentice and Proficient. Biserial correlation between item score and student performance level, with performance level dichotomized as either Apprentice or below or Proficient or above.
6. D_{P-D} biserial correlation	Discrimination between Proficient and Distinguished. Biserial correlation between item score and student performance level, with performance level dichotomized as either Proficient or below or Distinguished.
7. DIF-SMD Male-Female	Dorans and Schmitt (1991)* standardized mean difference for differential item functioning, comparing groups of male and female students matched on total common-item scores. A positive DIF-SMD score indicates relative to other items, this item favors males. A negative DIF-SMD indicates relative to other items, this item favors females.
8. DIF-SMD White-Black	Standardized mean difference for differential item functioning, comparing groups of White and Black students matched on total common-item scores. A positive DIF-SMD score indicates relative to other items, this item favors White students. A negative DIF-SMD indicates relative to other items, this item favors Black students.
9. Coefficient Alpha	Internal consistency statistic calculated for common items and for each subform calculated for total set of items within the appropriate content category.

*Dorans, N., & Schmitt, A.P. (1991). *Constructed-response and differential item functioning: A pragmatic approach (ETS research report No. 91-47)*. Princeton, NJ: Educational Testing Service.

TABLE 4-5
SAMPLE OPEN-RESPONSE ITEM ANALYSIS OUTPUT

				Frequency distribution				
Form	N	Mean	Std. Dev	0	1	2	3	4
Overall	49704	2.43	.837	1293	2857	23042	17963	4549
1	4174	2.29	.874	118	494	1915	1336	311
2	4139	2.46	.836	111	241	1797	1631	359
3	4152	2.44	.804	99	181	2030	1489	353
4	4165	2.47	.837	106	239	1806	1630	384
5	4158	2.42	.824	118	167	2123	1364	386
6	4175	2.37	.855	117	358	1891	1479	330
7	4136	2.49	.845	113	192	1806	1597	428
8	4160	2.47	.839	107	160	2020	1408	465
9	4142	2.47	.830	117	214	1773	1691	347
10	4120	2.46	.830	87	144	2063	1459	367
11	4088	2.42	.844	96	277	1906	1417	392
12	4095	2.47	.838	104	190	1912	1462	427

TABLE 4-5, continued
SAMPLE OPEN-RESPONSE ITEM ANALYSIS OUTPUT

Form	Item-Total Corr.	N-A biserial	A-P Biserial	P-D Biserial	Smd -Gender	Smd-Race
Overall	.513	.486	.376	.130	.027	.009
1	.568	.534	.435	.175	•	•
2	.525	.515	.356	.116	•	•
3	.501	.478	.352	.125	•	•
4	.516	.499	.363	.116	•	•
5	.448	.429	.327	.134	•	•
6	.552	.510	.406	.171	•	•
7	.503	.486	.368	.103	•	•
8	.481	.445	.398	.103	•	•
9	.504	.488	.364	.118	•	•
10	.484	.446	.371	.111	•	•
11	.552	.500	.412	.177	•	•
12	.511	.483	.374	.104	•	•

Note: The SMD statistic was not calculated by form for common items.

4. **DEVELOP CONCEPTS FOR NEW ITEMS.** The Content Advisory Committees review the existing item pool for areas that need to be replenished. For identified areas, the appropriate Content Advisory Committee brainstorms outlines for several possible items.

5. **WRITE INITIAL DRAFT OF NEW ITEMS.** After the first meeting, ASME staff (typically the test developer assigned to that committee) take the ideas generated by the Content Advisory Committees and develop initial drafts of new items.

6. **REVIEW AND REVISE INITIAL ITEM DRAFTS.** The initial item drafts are reviewed by ASME test development supervisors. Potential improvements are discussed with the item writers.

7. **REVIEW AND SELECTION OF ITEMS FOR FIELD TESTING BY CONTENT ADVISORY COMMITTEES.** At their second meeting, the Content Advisory Committees review, suggest revisions, and select items for field testing.

8. **ASSEMBLE FIELD TEST BOOKLETS.** A test assembler (typically the test developer assigned to the pertinent committee) selects the items that will constitute each field test form.

9. **REVIEW FIELD TEST BOOKLETS.** The assembled field test booklet is reviewed by the appropriate ASME test development supervisor.

10. ADMINISTER FIELD TESTS. New questions are field tested, gathering 50 to 100 student responses per question (fewer for performance events), primarily using students from schools in California, Maine and Massachusetts.[2]

11. ANALYZE FIELD TEST RESULTS. After field-testing, student answers to open-response questions and to performance events are scored by ASME staff using preliminary holistic scoring guides. The development staff then prepare summaries of results, including comments and observations of scorers. Multiple-choice data are analyzed to produce item statistics used to identify potential problems with items. Questions are revised based on field test results prior to and during committee meetings as the actual student field-test responses are reviewed by committees.

12. SELECT NEW ITEMS FOR UPCOMING YEAR'S TEST FORMS. At their third meeting the Content Advisory Committees select the remaining test items.

13. REVIEW OF SELECTED ITEMS BY BIAS REVIEW COMMITTEE. After final item sets are selected, all items are submitted to the Bias Review Committee. The role of this committee is to ensure that the content of the test is fair and equitable for all students, and that the tests contain no material that could be construed to be racially-, ethnically-, or gender-stereotyped or biased toward any group. Each testing year this committee meets to review all open-response, multiple-choice, and performance event items for every grade level. Committee members use response forms to record their specific concerns about each item, and to provide suggestions for revision.

14. REVIEW OF SELECTED ITEMS BY ASME STAFF. Content Advisory and Bias Review Committees[3] concerns are reviewed, and any appropriate revisions are made.

15. REVIEW OF SELECTED ITEMS BY KDE STAFF. Items are reviewed and any appropriate revisions are negotiated with ASME.

16. REVISE SELECTED ITEMS AND ASSEMBLE TEST FORMS. ASME staff build the 12 test forms for each grade level. Performance event sets are finalized.

17. REVIEW OF EACH ASSEMBLED FORM FOR CONTENT AND SENSITIVITY BALANCE BY ASME STAFF. While each individual item has been reviewed, the mix of items on each form is now assessed to ensure that content, academic expectations, and politically and socially sensitive issues are in balance within each test form.

18. REVIEW OF EACH ASSEMBLED FORM FOR CONTENT AND SENSITIVITY BALANCE BY KDE STAFF. KDE staff review the final draft forms before signing off for forms to be typeset.

19. REVIEW OF FINAL TYPESET COPY FOR POTENTIAL TYPOGRAPHICAL ERRORS. ASME and KDE staff perform final review on typeset copy before giving approval to print.

---

[2] Starting with the 1994-95 tests, some item tryout will occur in a separate pretest section of the operational test, rather than in a field test. This will allow larger samples of data based on Kentucky students to be the basis of item analysis.

[3] Starting with the 1994-95 tests, some item tryout will occur in a separate pretest of the operational test, rather than in a field test. This will allow larger samples of data based on Kentucky students to be the basis of item anayysis.

SPECIAL STUDIES

In the spirit of continuous quality improvement, KDE has conducted several special studies as part of on-going test development. A summary of the results of those studies follows.

TEST SENSITIVITY REVIEW BY AD HOC PARENTS AND PUBLIC COMMITTEE. This committee met from June 8-10, 1994 to review all forms of the 1993-94 KIRIS accountability-grade tests for balance and sensitivity to different points of view. The committee consisted of 22 members who were nominated by the Kentucky Council of Churches, Kentucky Education Association, Kentucky Parents Teachers Association, Kentucky School for the Deaf, KIRIS Content Advisory Committees, and citizens who had previously expressed concerns about these issues.

A subset of six individuals expressed strong concerns that (1) the selection of test questions favored a liberal viewpoint, (2) many questions appear to seek a politically correct response, (3) there is a lack of patriotic and traditional literature used in prompts, (4) religion is often cast in a negative light, (5) multiculturalism is overemphasized, (6) environmental issues are overemphasized because they are addressed in Science, Vocational Studies/Practical Living, and Social Studies, (7) the feminist viewpoint is given preference over traditionalist views.

Upon discussion, the consensus of the committee was that only one of these concerns – overemphasis of environmental issues – was well founded. Eighteen percent of the open response items (21 out of 116 items) were either about the environment or were set in a context related to the environment. Within the three content areas where these items appeared (Science, Vocational Studies/Practical Living, and Social Studies) the percent was considerably higher. In addition, it appears a disproportionate number of passages for the reading test were somber, bordering on depressing.

Based on these findings, additional balance and sensitivity steps and checks were added to the test development process.

ANALYSIS OF CONTENT COVERAGE BY NATIONAL EXPERTS. Content coverage of the multiple choice and open-response tests in reading, mathematics, social studies and science for 1993-94 was evaluated following administration of the these tests. For each content area, a review by KDE content area specialists and two national curriculum experts indicated complete content coverage when considered at the school-level, with multiple-choice and open-response items covering every category in the NAEP-based content frameworks. When considered at the student level, multiple-choice items also covered every content category, while open-response items covered all but five of the ninety content framework categories: ***Number*** in Grade 12 mathematics; ***Data Gathering and Reduction*** in grade 12 science; and, ***Astronomy, Matter, and Geology*** and ***Natural Resources*** in grade 4 science. (For additional information, see ***Further Consideration of Issues Related to the Inclusion of Multiple-Choice Items in KIRIS Reported Scores***), available from KDE.

AD HOC COMMITTEE REVIEW OF KIRIS APPROPRIATENESS FOR DEAF AND HEARING IMPAIRED STUDENTS. In October 1994 a committee which included deaf students and parents, teachers, and administrators of deaf and hearing impaired students met to review the KIRIS assessments and make recommendations about the proportion of items that (1) were fully appropriate for deaf students, (2) were appropriate for deaf or hearing impaired students, but for which the difficulty of the item might be affected by factors other than academic achievement in the subject area being assessed, and (3) were not at all appropriate for deaf or severely hearing impaired students. After this review, the committee discussed test administration and appropriate accommodations. Following are some key points from the meeting.

- The majority of the committee felt strongly that although the assessment of reading and writing skills was critically important, having all subject areas so dependent on this form of communication posed a particular challenge and perhaps a barrier to success for deaf and severely hearing impaired students.

- Several items were completely inappropriate for deaf students.
- Several modifications related to vocabulary and syntactic construction were suggested for particular questions. Further discussion led to the suggestion that a video American Sign Language interpretation might be the most appropriate single accommodation to consider making available.

DEVELOPMENT OF SCORING GUIDES

Scoring guides and training materials for open-response questions that were carried over from 1991-92 testing remained intact to ensure comparability of scores. New training materials were developed by test development staff and Advanced Systems' scoring staffs for those items new to the assessment in 1992-93 and 1993-94.

During initial test development meetings, the advisory committees discussed and drafted scoring guides for open-response questions. These guides were finalized by Advanced Systems' staff and scorers and reviewed by Department of Education content specialists. Scoring guides were then compared to actual papers returned from testing, measuring the expectations for responses against real student work. After this review, scoring guides were revised and refined yet again. Finally, student papers were selected for training packs. Typically, two sets of training packs per question were assembled. One contained clear exemplars of student work at each score point and was used to orient scorers to the range of responses and associated score points for each question. The second pack contained less clear-cut examples of score points and was used in training to gauge readers' internalization of the score points and readiness to begin scoring.

CHAPTER 5
SETTING PERFORMANCE STANDARDS

There are at least two meanings to "*standards*." The first is content standards: the decision about the range of knowledge and skills for which students will be held accountable. The other is performance standards: how well students will be expected to perform. Issues of content standards were resolved early in the project when the content area advisory committees made decisions about the content of the tests. Those issues and decisions are discussed in Chapter 3 of this report. This chapter will review the procedures that went into developing the performance standards.

STANDARD SETTING FROM A SCHOOL-LEVEL ACCOUNTABILITY PERSPECTIVE

The psychometric literature on standard setting deals exclusively with setting standards that are applied to individuals, typically in a high-stakes environment for those individuals. The KERA accountability system is somewhat different, as its primary purpose is setting standards upon which schools will be held accountable. Operationally, this translates into a policy-based assignment of points to each performance level and a requirement for schools to make continual progress on this weighted index.[1]

SELECTING THE NUMBER OF STANDARDS

Given the language of KERA, it was clear from the beginning of the project that some form of standard setting would be necessary. KERA calls for determining the "*proportion of successful students*" in a school; this phrase alone provides a mandate for standard setting. Without standards, a judgment about whether a student is successful could not be made.

---

[1] The weighted index is described in chapter 10.

Technically speaking, determining the proportion of successful students calls for one performance standard to be set. Those performing above the standard are successful; those below are not. In its proposal, Advanced Systems pointed out the negative effects of systems operating with just one standard. When other states have done that, usually in association with minimum competency programs, the standards have been low, so that almost all students have a reasonable prospect, with moderate effort, to achieve them. Because many students can reach the standards with minimal effort, the assessment program provides them with no motivation for improvement. Such a system clearly does not reflect KERA's main thrust. If, on the other hand, a single standard were set at the high level called for by KERA, many students, at least currently, would find the standard so high that they could not expect to achieve it even with considerable effort. This would have a negative effect on the effort that students put into their schoolwork and the assessment.

It was clear, therefore, that more than one standard was necessary; it was also important to keep standards to a reasonable number[2]. Too many standards produce at least two negative consequences. First, the differences between students performing at different standards is trivial, and classification can become quite arbitrary. Second, it is important that everyone involved in reviewing and interpreting reports, including parents, understand the standards. A large number of standards cannot be effectively communicated to diverse audiences.

Given these considerations, it was clear that four to six levels (three to five standards) would be needed. Six was rejected as too many. Systems with six levels are generally regarded as having three levels of success (high, higher, and highest) and three levels of non-success. Given KERA's belief that all students should be able to reach a high level of performance, the project staff felt that three levels of success were too many. They felt there should be no more than two levels of success, and so proposed a system with five levels. The original scoring guides for writing portfolios included five levels derived from this orientation. As the year progressed, however, it became apparent that five levels was also a poor choice for at least two reasons. First, because most grade scales used in schools have five levels (A, B, C, D, F), the relationship between grades and the assessment standards was sure to be confused. Additionally, it was felt that two levels of non-success were sufficient: that students below the standard will need to know only whether they are approaching the standard or still quite removed from it. Having three levels of non-success and only two levels of success places undue emphasis on the negative. By a process of elimination, therefore, a system specifying four levels was established.

After this decision was made, some concern was raised about that number by people who were aware of the experiences of other states that had used similar scales. They found that scoring was considerably less reliable with four levels, because judges tended to rate all students as being at either the second or third levels; there was also considerable confusion about which students were at each level. Kentucky's four-point system is considerably different, however, from the four-point scales used in other states. The third point on Kentucky's scale has been considered, from the beginning, to be the "high level of achievement" referenced by KERA, not performance that currently would be considered average. It was expected that most students would perform initially at the first or second level, with the upper two points in the scoring guides reserved for high and very high quality work. The difficulties encountered by other states using a four-point scale, where the third point simply meant "somewhat above average," were not likely to be faced by Kentucky, at least in the first years of the program.

The use of four levels of success remains controversial. Four levels produces less differentiation among students than many teachers find desirable. Some educators believe additional low levels would allow schools to be rewarded for progress with students who started with a very low level of achievement. This is particularly problematic with portfolios, where the number of score points and performance categories are equal. For example, the Writing Advisory Committee is concerned that their is no incentive for teachers to have low achieving students complete their portfolios.

---

[2] Psychometrically there would have been advantage to using a more continuous scale rather than discrete standard categories. However, at the time, this was viewed as inconsistent with the law.

NAMING THE STANDARDS

Naming the levels was another process involving considerable time and thought. Several possibilities were raised. One issue concerned naming the third (successful) level. It was first labeled "world-class," because some of the rhetoric associated with KERA used that term (although it is not used in KERA itself). "World-class" was discarded, however, because of the confusion, rather than clarification, it offered. People thought they knew what "world-class" meant, and presumed that others were using the same definition. Often, we found, they did not, and we felt we should find a term that was more specific. To some extent, "competent" would have been a good choice because it directly reflected the meaning of success as defined by KERA. But "competent" has a connotation of pedantic — satisfactory, but not necessarily well-grounded — that did not match the prestige to be associated with students who had reached the high level of success KERA required. That term also was rejected. "*Proficient*" evolved as the popular choice, because it conveyed solid competence. Its only drawback was that NAEP uses that term; since KIRIS will be linked to NAEP, and because NAEP's standard of "proficient" likely will be at least somewhat different from Kentucky's, there was concern about confusion between the two. However, all things considered, "*Proficient*" was judged to be the most appropriate term.

Defining the other levels was equally problematic. At first, the lowest level was called "unsatisfactory." It became clear, however, that the negative connotations associated with that term rendered it unacceptable. Other terms such as "minimally competent" also carried negative connotations, the most disturbing of which was the possibility of a long-term classification of the student as a failure. A term was needed that connoted a low level of performance, but also the expectation was that this level was transitory; that the student would grow to higher levels in the future. "Novice" came to be used as more people considered the issues. "*Apprentice*," the term used to describe the second level, became a natural extension of "*Novice*."

As much thought went into the name for the highest level. It was intended that the term convey exceptional performance — a level of performance beyond that established by KERA, that few students might attain. "*Advanced*" was well received at first, but was unacceptable to members of the twelfth-grade standard-setting committee. They felt that the best work was not "advanced," just more detailed and better organized than proficient work. Eventually, in part because of its academic ring, "*Distinguished*" became the label of choice for the highest level.

STANDARD SETTING PROCESS – GENERAL ISSUES

There are two types of standard setting processes: contrasting groups and expert judgment (Zieky & Livingston, 1977)[3], but ultimately all standard setting methods rest on subjective opinion (Glass, 1978)[4]. The challenge in standard setting is to minimize the subjectivity by providing consistent supporting information to those involved in the standard setting process.

Performance standards for KIRIS tests in mathematics, reading, science, social studies, and writing were set in 1992. In June, 1992, scoring standards for writing were adjusted to be consistent with the four categories reported for other subject areas. In 1993, standard setting was performed for arts and humanities, practical living/vocational studies, mathematics portfolio, and alternate portfolio.

---

[3] Zieky, M.J., & Livingston, S.A. (1977). *Manual for setting standards on the Basic Skills Assessment Tests.* Princeton, NJ: Educational Testing Service.

[4] Glass, G.V. (1978). Standards and criteria. *Journal of Educational Measurement, 15,* 237-262.

IMPACT OF A LIMITED NUMBER OF SCORE POINTS. The typical way subjectivity is minimized in the standard setting process is by aggregating expert judgments over a large number of items. A typical test used to make high stakes decisions about individuals has 60-200 items. Thus, for any random errors in judgment about an item, there is much opportunity for a corresponding opposite error, and overall, random judgment errors are minimized.

In 1991-92, there were three common items and 13 score-scale points (each item was scored on a 0-4 scale). Since 13 scale points were being mapped to four performance categories, a difference of one point in a category boundary could have a large effect on the proportion of people in each category (particularly so for the novice-apprentice boundary).

SETTING STANDARDS BASED ON OPEN-RESPONSE ITEMS

The open-response tests consisted of common and matrix-sampled questions. Standards were set on the common items first; standards for the matrix-sampled questions were derived from the data describing the relationship between the two types of questions.

KIRIS standard setting required using some strategies that are different from those used by testing programs relying either exclusively or primarily on multiple-choice questions and reporting their results in terms of scaled scores. First, standard setting was done on performance (open-response) items only; no multiple-choice questions were included in the process. Second, standard setting was done by examining actual student work rather than scoring guides or other intermediate elements of the scoring process. Finally, standard setting was done either using a conjunctive model, that is, on the basis of all work by a student in a content area (mathematics, reading, and social studies), or using a compensatory model, that is, by summing standards based on individual items (science, arts and humanities, and practical living/vocational studies).

COMMITTEE MEMBERSHIP. Committees were selected to take part in the standard-setting process. These committees consisted of staff from two divisions (KIRIS and Curriculum Development) within the Department of Education, and members of the content advisory committees for test development. The content advisory committees included people with a range of backgrounds, the majority of them classroom teachers. Tables 5-1 and 5-2 describe each standard setting committee in terms of member experience and demographics, respectively.

TABLE 5-1
STANDARD SETTING COMMITTEES - MEMBER EXPERIENCE

Content Area	Teachers	Administrators	KDE Staff	Others
Mathematics	11	2	2	0
Reading	7	1	2	3
Science	9	0	1	2
Social Studies	7	3	3	0
Writing	37	5	3	11
Arts and Humanities	9	1	1	2
Practical Living/ Vocational Studies	13	3	4	5

TABLE 5-2
STANDARD SETTING COMMITTEES - MEMBER DEMOGRAPHICS[5]

Content Area	Male	Female	Minority	Non-Minority	Region							
					1	2	3	4	5	6	7	8
Mathematics	5	10	1	14	0	1	1	2	5	1	2	1
Reading	0	14	0	14	0	1	1	4	2	1	0	0
Science	3	8	1	10	3	0	2	2	3	1	0	0
Social Studies	7	6	0	13	1	1	6	3	0	0	0	1
Writing	5	37	0	42	5	4	7	3	10	5	4	4
Arts and Humanities	6	7	2	11	1	0	4	1	4	3	0	0
Practical Living/ Vocational Studies	8	14	2	20	3	3	3	3	6	2	2	0

SCORING GUIDES. ASME, with substantial input from KDE and the content area advisory committees, developed scoring guides for each open-response question in reading, mathematics, science and social studies. Each question was scored on a scale of 0-4. (At grade 12, papers were scored on a scale of 1-5, and then converted to the 0-4 scale by subtracting one from each score). While detailed, specific scoring guides were developed for each question, the general pattern of assigning scores to papers is described in Table 5-3.

---

[5] Due to lost records, different sets of mutually exclusive categories may not add up to the total.

TABLE 5-3
GENERAL ITEM SCORING GUIDE

Score	Description
0	No response or irrelevant response
1	Minimal or incorrect response, demonstrating little or no understanding of relevant concepts or processes
2	Partially correct/partially complete response, without support, demonstrating gaps in conceptual understanding
3	Complete, reasonably supported response, demonstrating understanding of major concepts and processes
4	Complete, insightful, strongly supported response, demonstrating in-depth understanding of relevant concepts and processes

STANDARD SETTING PROCESS. After scoring was complete, the standard-setting committees were convened to establish a protocol for translating scores to standards. Because the scoring guides and standards for writing had already been disseminated and it was important that the standards across the content areas be comparable, the standard-setting committees were first trained on the writing scoring guides. After training, they were provided with several examples of scored writing samples.

After the committee members were satisfied that they understood the standards set in writing, each participant was trained in the scoring guides for the open-response questions in their content area. Once they were sufficiently trained to be competent scorers, they were given several papers that had been assigned each score point on the scale described above. For each question, they independently decided which score reflected each level of performance. After reaching individual conclusions, they debated their decisions as a group. After about an hour, they reconsidered their original decisions and altered them as they saw fit. At this point, there was a measure of consensus about the correspondence between score points and performance levels. Committee members felt that scores of 0 and 1 generally indicated *Novice* work, 2 was *Apprentice*, 3 was *Proficient*, and 4 was *Distinguished*. There were several exceptions to this general pattern, however, as well as a feeling that the scoring guides, as created, did not always require distinguished-level work to obtain even the highest scores. This is not surprising, because of items and their scoring guides were not constructed to map to the student performance categories. Moreover,because items and scoring guides were so constructed, this would limit the ability to develop items to maximally differentiate at different performance boundaries.

It was also not uncommon for committee members to disagree about the level of performance exemplified by different papers. Thus, for example, some felt that a score of 3 on one question reflected *Apprentice*, not *Proficient*, work, while others felt that the pattern applied to another question but not that one. Many of these disagreements disappeared, however, when committee members examined a student's performance across all three questions in the subject area. Once that was done, the committees came to fairly rapid agreement on rules for translating scores into proficiency levels — for example, "If the student scores a 4 on one question and 3's on the other two, that student is proficient, regardless of which question received the score of 4." Table 5-4 demonstrates the extent to which committee members were consistent in their judgements.

TABLE 5-4
CONSISTENCY OF COMMITTEE MEMBER STANDARD-SETTING JUDGEMENTS – GRADE 12 READING
NUMBER OF COMMITTEE MEMBERS SELECTING EACH SCORE POINT FOR EACH PROFICIENCY CATEGORY FOR EACH OF THREE COMMON ITEMS

Item Score	Item 2				Item 3				Item 4			
	N	A	P	D	N	A	P	D	N	A	P	D
5			6	1			3	4			3	4
4		6	1			4	3			3	4	
3	1	6			1	6				7		
2	7				7				7			
1	7				7				7			

Table 5-4 shows that for item 2, all 7 judges thought a score of 1 or a score of 2 indicated the student was a *Novice*. One judge thought a 3 response was also indicative of a *Novice* student, but the other 6 judges thought a score of 3 was an *Apprentice* level response. The seven judges were very consistent in their rating of items. The correlation of item score with proficiency level (with proficiency level coded 1-4) across the seven judges was .91, .90, and .92 for the three grade-12 reading common items.

After the preliminary decision rules were made, the committees were given several papers that met the upper and lower criteria for each score point; they discussed whether the paper at the lower point satisfied their criteria for minimal performance at the designated performance level, and whether the paper at the upper point fit that level and not the next higher one. We found that the committees did not change their judgments; in fact, reviewing these papers seemed to reinforce their opinions about the accuracy of their preliminary decision rules.

As a final activity, the committees were given information about the percentage of students attaining each performance level. Generally, very small percentages of students achieved scores that would categorize them as Proficient or Distinguished. The committees were allowed to reevaluate their decision rules in light of the data. In only one case out of twelve (grade 4 Reading) did the committee wish to reconsider their judgments, and even in that case (in part through a miscommunication of the directions) did not change their decision rules.

STANDARD SETTING RESULTS. This process established the relationship between scores and performance levels for the three common items for each content area. Those relationships are provided in Table 5-5 - 5-8, for mathematics, reading, science, and social studies, respectively.

TABLE 5-5
PERFORMANCE STANDARDS FOR 1991-92 OPEN-RESPONSE MATHEMATICS TESTS

Grade	Proficiency Level	Requirements for Performance Level
4	Distinguished	12
	Proficient	9-11, with no score less than 3
	Apprentice	6-8
	Novice	0-5
8	Distinguished	12
	Proficient	9-11, with no score of 1 or 2
	Apprentice	7-10, with no combination of 1s and 2s
	Novice	0- 7 unless Apprentice
12	Distinguished	14 or 15, but must have a 5 on question 2
	Proficient	12-14, unless Distinguished
	Apprentice	8-11
	Novice	3-7

TABLE 5-6
PERFORMANCE STANDARDS FOR 1991-92 READING OPEN-RESPONSE TESTS

Grade	Proficiency Level	Requirements for Performance Level
4	Distinguished	12
	Proficient	9-11
	Apprentice	5-8
	Novice	0-4
8	Distinguished	12
	Proficient	8-11, Must have a score of at least 3 on questions 1 & 3 and no less than 2 on question 2
	Apprentice	5-8, unless Proficient, with no more than one score less than 2
	Novice	0-4
12	Distinguished	15
	Proficient	At least 12 points, and nothing less than a 3
	Apprentice	8-12 points, with no more than one score less than 3
	Novice	3-8, 2 or more scores less than 3

TABLE 5-7
PERFORMANCE STANDARDS FOR 1991-92 SCIENCE OPEN-RESPONSE TESTS

Grade	Proficiency Level	Requirements for Performance Level
4	Distinguished	12
	Proficient	10-11
	Apprentice	6-9
	Novice	0-5
8	Distinguished	12
	Proficient	10-11
	Apprentice	5-9
	Novice	0-4
12	Distinguished	15
	Proficient	13-14
	Apprentice	8-12
	Novice	3-7

TABLE 5-8
PERFORMANCE STANDARDS FOR 1991-92 SOCIAL STUDIES OPEN-RESPONSE TESTS

Grade	Proficiency Level	Requirements for Performance Level
4	Distinguished	11-12
	Proficient	8-10, with no more than one 2 on any question and no less than 2 on any question
	Apprentice	5-8, with no more than one score less than 2
	Novice	0-4
8	Distinguished	11-12
	Proficient	8-10, with a score of 8 requiring item scores of 3,3, and 2
	Apprentice	5-8, if not Proficient, and with no more than one item score of 1
	Novice	0-5, if not Apprentice
12	Distinguished	15, or 14 if a 4 on question 2
	Proficient	11 (if a 3 on question 2), 12-13, or 14 (if a 4 on either questions 3 or 4)
	Apprentice	8-11, unless Proficient
	Novice	3-7

An important result of the process is that all standards were set on the basis of actual student performance — not on the basis of presumed skill in selecting correct responses to multiple-choice questions, and not on the standard-setting guides. One measure of the effectiveness of any standard-setting procedure is on how the decisions of the committee are accepted. The project staff was very comfortable with the final decisions of the standard-setting committees. In addition, when the standards were communicated statewide as part of reporting, it was clear that people across the state found the standards credible. There was little questioning that the standards were high, but fair.

MAINTAINING COMMON ITEM STANDARDS ACROSS YEARS

Standards were maintained across years by equating. Using Samejima's two-parameter graded-response model (1969), and estimating item and person parameters with MULTILOG® (1991), the theta that cut off an equal percent of examinees as were in each performance category was used to maintain cut scores in 1992-93 and 1993-94, with two exceptions:

1. No reversals were allowed across forms; that is, if a raw score of 6 were *Proficient* on one form of the mathematics test, a raw score of 7 could not be *Apprentice* on another form; (this does not effect matrix-based scores), and

2. "Doglegging" (a procedure used in the equating of SAT forms) was used to ensure that students could attain a score of *Distinguished* regardless of which form they took; that is, on policy grounds, a perfect score was declared *Distinguished*, even if the equating would have declared that score a *Proficient*[6].

More information about equating can be found in Chapter 9.

EFFECT OF SWITCHING FROM CONJUNCTIVE TO COMPENSATORY STANDARDS

With the exception of Science, standards were set in 1991-92 using a conjunctive model (i.e., to reach a particular standard required more than just a set number of points[7]). Starting in 1992-93, all standards were based on a compensatory model (i.e., fewer points on one item could be compensated for by more points on another item). While the two standard setting models were equated to maintain the same proportion of people in each proficiency category (with the two policy-based exceptions mentioned before), the particular students achieving each standard would have differed.

EXTENDING THE STANDARDS TO MATRIX ITEMS

Standards for the matrix-sampled questions were derived from those established on the common items, using a regression approach. For each of the 12 open-response items in a content area, an estimate of the performance level corresponding to each open-response score was computed using the following procedure:

a. Based on the percent of students within each performance level on the common items, a standardized score, z, was assigned to each performance level to represent the percent of students falling within or above that performance level in a standardized normal distribution. For example, suppose that 10 percent of the students in a content area were classified as *Distinguished*, 15 percent as *Proficient*, 35 percent as *Apprentice*, and 40 percent as *Novice*. The *Proficient* level would contain the highest-scoring 76 to 90 percent of the students. The midpoint of this interval is 83 percent. The z-score that cuts off the bottom 83 percent of a normal distribution is 0.96. Thus, for this example, the z-score that would be assigned to the *Proficient* level is 0.96.

---

[6] In most cases, doglegging did not significantly effect the proportion of students achieving any proficiency level. Often the increase in percent of Distinguished students was less than .1%.

[7] For example, to achieve the Apprentice standard in 4th grade social studies required 5-8 points, with no more than one score less than 2.

b. The standard scores representing each performance level were then used as the dependent variable in a regression procedure to predict a students' performance level based on performance on a matrix-sampled item. The independent variables in the regression procedure were dummy-coded variables representing each of the five possible open-response scores. A student was assigned a 1 on the variable representing the score received and a 0 on the remaining four variables.

c. Regression equations were computed for each of the 12 matrix-sampled items, and then predicted performance level scores were calculated for each possible open-response score. This resulted in a total of 60 predicted performance level scores — five scores for each of 12 items.

d. The 60 predicted scores were then rank ordered and the number of students achieving each score was noted. Cutoffs for each performance level were then determined so that the percent of students classified at each performance level on the matrix-sampled items was as equal as possible to the percent of students classified at each level on the common items.

e. When the cutoff points were established, each student was then assigned to a performance level based on his/her score on the matrix-sampled item.

Tables showing the final results of converting raw scores to performance levels is provided in the 1991-92 Technical Report.

MATRIX SCORES IN ARTS AND HUMANITIES, PRACTICAL LIVING, AND VOCATIONAL STUDIES

In 1991-92, KIRIS tested reading, mathematics, science, social studies, and writing. In 1992-93, testing in the arts and humanities, and practical living/vocational studies was added. In all subject areas, responses were scored on a point scale from zero to four. Due to the varying difficulty of items, numerical scores do not translate directly into the *Novice*, *Apprentice*, *Proficient*, and *Distinguished* performance levels used for reporting. Therefore, standards had to be set to determine performance levels, as was done in 1992 for the initial core subjects.

For the subject areas tested during 1991-92, students answered three open-response questions in each of four areas, and produced a writing portfolio. For 1992-93's new subject areas, approximately 11 interdisciplinary open-response questions were matrix-sampled throughout 12 forms at each grade level[8]. Most students answered one of the questions; some students answered two, and some answered none. Because most students produced only one response (rather than several measures) to be scored, a new method had to be devised to set the performance-level standards. However, to the extent possible, the new standard-setting process was based on the same operational definitions and reflected the same goals and outcomes as the standards for the other subject areas.

The procedure for setting performance-level standards for the new assessment areas began the same way as the standard-setting procedure in the other subject areas. A committee of experts, consisting chiefly of the advisory committees, first were trained in the methods for scoring and evaluating writing, and then were asked to set standards for their subject areas consistent with the operational definitions for writing.

---

[8] To minimize total testing time, no additional items were added for arts & humanities or practical living/vocational studies. Instead interdisciplinary items that measured one of the new content areas and one of the old (mathematics, reading, science, or social studies) were added in to the test. These items were scored using a single scoring guide and counted toward two scores. Interdisciplinary items are being phased out, beginning with the 1994-95 testing year.

Once the committee was thoroughly versed in the standards and standard-setting procedures for writing, its members were trained to score the open-response arts and humanities, practical living, and vocational studies questions. Understanding scoring is a prerequisite for setting performance standards; one must understand what is required for a response to a particular item to receive a score of 3, for example, before one can select a performance level for that score point.

Before performance-level standards could be set, the data were analyzed in order to place all the items and item responses on a common scale. This process balanced out differences in the level of difficulty or the stringency of scoring among the items. Also, scoring and test development staff collected six to eight student responses for each item at each score point so the panel could see the variety of actual student work at each score point.

Since each student completed only one item in these subject areas, performance on that one item was the basis for assignment to a performance level. (In all other subject areas, performance levels were assigned to the aggregated scores of all items in the subject.) Based on personal judgment, each committee member assigned one of the four performance levels to each item/score point combination. That is, looking at the quality of work represented by score-point samples, the judges decided if that work characterized a student of *Novice*, *Apprentice*, *Proficient*, or *Distinguished* caliber. They knew at the outset that some items were more difficult than others, so that they might judge a 2 on one item as *Proficient*, for example, while deeming that a 3 on another represented *Novice* performance.

Certain item/score point combinations had obvious agreement; that is, everyone on the committee assigned the same performance level to that combination. When that happened, the standard for that item/score point combination was clear. Substantial (one-third or more) disagreement about item/score point combinations, however, indicated where cut points between the performance levels would likely fall. In other words, if half of the evaluations placed a certain item/score point combination at *Novice*, and the other half placed it at *Apprentice*, the number representing that item/score point combination on the scaled score became the cut point. All item/score point combinations below that cut point were classified as *Novice*, and all above the cut point were *Apprentice* (up to the next determined cut point).

RESULTS. When both subject areas at all three grade levels were scaled out in this way, five of the six scales had virtually identical cut points, and the sixth (grade 4 practical living and vocational studies) had only a minor variation. Note that practical living and vocational studies are combined (PL/VS) for standard-setting and reporting purposes. Tables 5-9 and 5-10 present the results for these subject areas.

TABLE 5-9
PERFORMANCE STANDARDS FOR 1992-93 PRACTICAL LIVING/VOCATIONAL STUDIES OPEN-RESPONSE TESTS REQUIREMENTS FOR PERFORMANCE LEVEL, BY ITEM

Grade	Proficiency Level	Item										
		1	2	3	4	5	6	7	8	9	10	11
4	Distinguished		4			4	4	4	4		4	
	Proficient	4		4	4			3			3	4
	Apprenice	3	3	3	3	3	3		3	4		3
	Novice	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2,3	1,2	1,2
8	Distinguished								4			
	Proficient	4	4	4	4	3,4	4	4	3	4	3,4	3,4
	Apprentice	3	3	3	3	2	3	3	2	3		2
	Novice	1,2	1,2	1,2	1,2	1	1,2	1,2	1	1,2	1,2	1
12	Distinguished		4									
	Proficient	4		4	3,4	4	4	4	3,4	4		
	Apprentice	3	3	3		3	3	3		3		
	Novice	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2		

TABLE 5-10
PERFORMANCE STANDARDS FOR 1992-93 ARTS & HUMANITIES OPEN-RESPONSE TESTS
REQUIREMENTS FOR PERFORMANCE LEVEL, BY ITEM

Grade	Proficiency Level	Item								
		1	2	3	4	5	6	7	8	9
4	Distinguished									
	Proficient	4	4	4	4	4	4	4	4	4
	Apprentice	3	3	3	3	3	3	3	3	3
	Novice	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2
8	Distinguished					4				
	Proficient	4	4	4	3,4	3	4			
	Apprentice	3	3			2				
	Novice	1,2	1,2	1,2,3	1,2	1	1,2,3			
12	Distinguished			4			3,4			
	Proficient	4	4	3	4	4	2	4	4	4
	Apprentice	3	3		2,3	3		3	3	3
	Novice	1,2	1,2	1,2	1	1,2	1	1,2	1,2	1,2

EXTENDING THE RESULTS TO PERFORMANCE EVENTS

For a variety of reasons, including the availability of experienced people and limited familiarity with students' mastery of performance events, it was decided not to conduct a separate standard-setting activity for these items. Nevertheless, to make the performance events reporting consistent with the remainder of the test results, it was necessary to convert the performance events scores into standards. In order to accomplish this within the resources of the program, a simplifying assumption was made that the percentage of students at each proficiency level on the performance events was approximately the same as that of the open-response questions. Thus, just as for the open-response matrix-sampled questions, the data from the common open-response questions was used as the criterion of success.

In order to accomplish this, a procedure similar to the one described above for the open-response matrix-sampled questions was used, except that no regression was done. For each content area and each grade level, the percentage of students at the *Distinguished* level was obtained. This was compared to the frequency distributions for each performance event within the content area. To obtain the number of students taking events in that content area that were to be classified as *Distinguished*, the number of students taking performance events within the content area was determined and multiplied by the percentage of students to be labeled as *Distinguished*. The percentage of students obtaining the highest score on each event within the content area was then calculated. The score with the smallest percentage was assumed to be the most difficult score to obtain across those events. The number of students achieving that score was determined. If that number was greater than the number that should be classified as *Distinguished*, then only those students obtaining that score were classified as *Distinguished*. If not, then the next highest score/frequency combination was determined, and added to the first. This process continued until the number of students classified as *Distinguished* on the performance events approximately equalled the number so classified on the transitional tests.

Once the highest score/event combinations were assigned to the *Distinguished* level, the process continued to determine which score/event combinations should be assigned to the *Proficient* level, using the same procedures. Finally, the combinations to be assigned to *Apprentice* were determined. The remaining lowest score/event combinations became the *Novice* level.

SETTING STANDARDS BASED ON WRITING PORTFOLIOS

Standard setting for the writing portfolios was conducted in February 1992. After the content of the portfolios was determined, it was necessary to deliver scoring guides to teachers statewide. The writing assessment is different from other subject areas because the writing standards must match the limited number of points on the writing scoring scale. With the limited number of score points in writing, however, if 4 sets too high a standard, it may very well be that 3 is much too low – and no intermediate score point is available. Therefore, the Scoring Guides and the standards must have a one-to-one correspondence.

The standard setting was done by the Writing Advisory Committee. As noted in an earlier chapter, the members of this committee included approximately ten teachers per accountability grade level, staff from the Kentucky Writing Project, university faculty, and members of the Department of Education. The committee reviewed actual student portfolios, selected sample portfolios that exemplified the standards, and completed the Scoring Guide to make operational the standards they had envisioned for the program.

REVISION OF WRITING SCORING GUIDES. During the summer of 1992, the Writing Advisory Committee revised the KIRIS Annotated Holistic Scoring Guide for writing portfolios so that it would parallel the performance levels of scoring guides in other content areas. The initial score points 1 and 2 were consolidated in the Novice performance level, score point 3 became Apprentice, score point 4 changed to Proficient, and score point 5 became Distinguished. These changes are listed in Table 5-11.

TABLE 5-11
CONVERSION OF WRITING PORTFOLIO SCORE POINTS TO PERFORMANCE LEVELS

1991-92 Score	1992-93 Performance Level
1	Novice
2	Novice
3	Apprentice
4	Proficient
5	Distinguished

Though the scoring model changed, it is important to note that Kentucky standards have remained constant, as have the benchmarks at each of the grade levels. The language of the 1992-93 scoring guide was revised to make the distinction between each performance level more apparent. The scoring guide criteria were also revised to reinforce understanding among students and teachers that the students' writing must communicate ideas and information to a variety of audiences for a variety of purposes, and that pieces from content areas other than English/language arts are significant portfolio pieces. The Author's Self-Assessment/Conference Form was updated to parallel the changes in the Scoring Guide.

ADDITIONAL SCORING GUIDE CHANGES. In 1993-94, in response to suggestions from teachers and others, the Scoring Guide and training materials for writing portfolios were supplemented with additional information. In addition, minor wording changes were made. However, the Kentucky writing standards originally set by the Advisory Committee remained unchanged. The reason for providing the supplementary materials was to clarify the standards and improve the ease of using the Scoring Guide.

The Writing Advisory Committee made the following changes to the criterion descriptors on the Scoring Guide in response to questions and suggestions from the field:

- ***Purpose/Approach*** was changed to ***Purpose/Audience***.
- ***Surface Features*** was changed to ***Correctness***.
- ***Wording*** was changed to ***Language***.

The section of the Guide previously referred to as ***Analytic Annotations*** is now called ***Instructional Analysis***. While scorers were previously asked to give two annotations per portfolio, they now may mark as many instructional strengths as they see fit for each portfolio.

The Scoring Guide now contains a section on ***Complete/Incomplete Portfolios***. This information was previously located in the ***Teacher's Handbook***.

TRAINING MATERIAL CHANGES. The Writing Advisory Committee decided to create one universal training packet to be used by all scorers, rather than the previous multiple levels of training. In addition, rather than creating two documents as had been done in the past, the ***Teachers' Handbook*** and all scoring training materials were combined into one document. The following changes and additions were made to training materials at each grade level:

- 4th Grade - The Committee added 1 High Apprentice and 1 High Proficient portfolio to the scoring training materials.
- 8th Grade - The Committee added 1 High Apprentice and 1 High Proficient portfolio to the scoring training materials.
- 12th Grade - The Committee replaced the Proficient Exemplar portfolio and added 2 High Proficient portfolios to the scoring training materials.

The Writing Advisory Committee voted unanimously to make no changes to the Benchmarks for any grade level.

SETTING STANDARDS BASED ON MATHEMATICS PORTFOLIOS

BACKGROUND. Since mathematics portfolios were new to the KIRIS assessment in 1992-93, a benchmarking process was needed to define the standards Kentucky expects for each grade level. The Holistic Scoring Guide was used to identify samples (benchmarks) that would convey these performance standards to teachers. Benchmark portfolios are whole portfolios that exemplify the language of the Holistic Scoring Guide and solidly represent a particular performance level. Since the benchmarks were set using actual Kentucky student portfolios, the benchmarking process could not take place until spring 1993, when the student work was available. As a result, teachers could not receive training or score portfolios during the school year. They were given the summer and early fall of 1993 to attend training and score the portfolios, with initial scores due in October, 1993. Now that benchmarking and standard-setting have been completed, teachers will be able to score portfolios within the school year.

The benchmark committee included the Mathematics Portfolio Regional Coordinators, university faculty, leading mathematics educators in Kentucky, and members of the KDE and Advanced Systems staff.

PORTFOLIOS USED IN BENCHMARKING. The portfolios used for benchmarking came from Kentucky students. It was important to collect portfolios from across the state and from all performance levels. Therefore, schools were identified based on several demographic and distribution criteria including geography, rural/urban classification, independent/county affiliation, and performance on the mathematics portion of the transitional test. KDE contacted the principals of 79 schools (24 elementary schools, 30 middle schools, and 25 high schools) asking each to identify a teacher to provide 25-30 mathematics portfolios for use in establishing benchmarks and creating a scoring manual. Participating teachers agreed to complete portfolio development approximately six weeks earlier than the official completion date, and to include in their students' portfolios the two sample tasks disseminated with the "Teacher's Guide Addendum."

BENCHMARKING PROCESS. Committee members were asked to prepare themselves for the benchmarking meetings by reflecting on what constitutes "good mathematics," reviewing the Holistic Scoring Guide, and reading five portfolios selected by KDE and Advanced Systems consultants. The meetings began with a general discussion of holistic scoring. Next, each member of the committee was asked to rate the five selected portfolios as "good," "okay," or "minimal," based on personal judgment. The committee discussed each portfolio and reached consensus on these ratings. This activity allowed all committee members to experience holistic scoring.

Next, the committee reviewed the portfolios according to the language on the Holistic Scoring Guide. For example, portfolios initially grouped as "good" were reexamined using the *Proficient* level criteria on the Holistic Scoring Guide. The committee then reached consensus on a performance level rating for each of the five portfolios. With the standards established, the next two days were spent reading many portfolios and rating them according to the language on the Holistic Scoring Guide. Selected portfolios were reviewed and discussed by a subcommittee that had particularly extensive experience with mathematics portfolios. The subcommittee recommended five portfolios at each level that were brought to the entire committee. These portfolios were reviewed and discussed until all members came to a consensus regarding the scoring of the selected portfolios. Those portfolios that best exemplified the language on the Holistic Scoring Guide were selected for inclusion in the teachers' scoring manual.

KDE and Advanced Systems' staff then wrote rationales for each portfolio, according to the criteria on the Holistic Scoring Guide. The committee's next step was to reevaluate its scoring decisions and review the rationales. Upon final committee approval, the portfolios and their accompanying rationales were compiled by grade level and published in the *KIRIS Mathematics Portfolio Assessment Scoring Training Manuals* (Grade 4, Grade 8, and High School).

Each manual contained 14 portfolios: six benchmark portfolios (two each *Novice, Apprentice,* and *Proficient*), four training packet portfolios, and four reflective packet portfolios. Each type of portfolio served a distinct purpose. Benchmark portfolios provided examples that reflected the "middle" of a particular performance level. Training portfolios included work of more uneven quality and offered discussion opportunities during training. Reflective portfolios were included to give teachers another chance to practice scoring before scoring their own portfolios.

Although the committee found work at the *Distinguished* level, they did not find *Distinguished* portfolios that reflected the "middle" of the *Distinguished* performance level. Therefore, no *Distinguished* benchmarks were included in any scoring training manual. However, some of the examples of *Proficient* student work included suggestions for improvements that would have merited them a *Distinguished* rating. In both the fourth grade and high school manuals, one *Distinguished* portfolio was included in the training packet portfolios. In addition, the committee outlined a procedure for identifying *Distinguished* benchmark portfolios which could be included in future training materials. An addendum published late in 1993 provided additional *Distinguished* portfolios for fourth and eighth grades.

MATHEMATICS PORTFOLIO SCORING MANUALS. The State Board for Elementary and Secondary Education approved moving the Mathematics Portfolio for Grade 4 to Grade 5 in 1994-95. Due to this change, a new Grade 5 Mathematics Portfolio Scoring Training Manual is being developed and by summer 1995, will be disseminated to teachers through District Assessment Coordinators. The manual will provide examples of grade 5 portfolios to clarify the language of the Holistic Scoring Guide and will be used to train teachers to score grade 5 portfolios. The manual will include the following:

- Four sections - one for each performance level of *Novice, Apprentice, Proficient,* and *Distinguished.*
- Two or three portfolios to illustrate each of the performance levels.
- Each portfolio will be accompanied by a Rationale, Holistic Scoring Guide, and Core Concept Demonstrator sheet

The manuals for Grades 8 and 12 were updated during 1994-95 in response to teacher requests to clarify language and to provide additional examples of student portfolios. The second edition of the manual includes these revisions:

- The portfolios are organized into four sections -- one for each performance level Novice, Apprentice, Proficient and Distinguished.
- Each section contains two or three portfolios including benchmarks (portfolios that exemplify a particular scoring level), high-ends (portfolios that demonstrate the upper end of performance levels), and/or special considerations (portfolios that may be particularly difficult to score).
- The portfolios were selected from the original Scoring Training Manuals and Consensus Portfolios, Summer 1994 Review Study, or were identified during the year as good examples. A rationale, 1994-95 Holistic Scoring Guide, and Core Concepts Demonstrator sheet accompanies each portfolio.

SETTING STANDARDS BASED ON ALTERNATE PORTFOLIOS

Alternate portfolios were implemented during 1992-93. These portfolios are described in Chapter 3. Following each of the steps described above for benchmarking mathematics portfolios, the Alternate Portfolio Advisory Committee selected benchmark portfolios at each performance level in June, 1993. Members of the Kentucky Systems Change Project staff, those primarily responsible for the development and implementation of alternate portfolios, attended the mathematics portfolio benchmarking sessions in order to learn and apply the benchmarking procedures for their use.

The committee reviewed a sample of 150 portfolios collected from various schools and selected portfolios that exemplified the performance-level standards. Student samples were chosen for scoring manuals for each level (elementary, middle, and high school), and rationales were written to accompany them. The committee also finalized the scoring guide to make operational the standards they had set for alternate portfolios and discussed teacher training needs. The *Kentucky Alternate Portfolio Project Teacher's Guide* contains the Holistic Scoring Guide developed for these portfolios.

VALIDATION OF STANDARDS

Expert judgment methods of standard setting leave open the question, "If a very different group of experts tried to set similar standards would the results have been radically different." Two such natural experiments exist to help Kentucky gather information about the validity of their standards. The National Assessment of Educational Progress has independently set four levels of performance. No staff or committee members who worked on the setting of Kentucky standards worked on the setting of NAEP standards. (The state of Maine has also set four levels of standards. Unfortunately, much of the data involving Maine was not available at the time this Technical Manual went to press.) In 1994, NAEP reported the result of its 1992 trial state assessment in reading and mathematics at grades 4 and 8. Table 5-12 compares the percent of Kentucky students in each Kentucky standard category and in each NAEP category. The table shows a marked similarity among the results of standard setting processes, with Kentucky having more difficult standards the NAEP in almost every case.

TABLE 5-12
COMPARISON OF KIRIS AND NAEP STANDARDS
PERCENT OF 1991-92 KENTUCKY STUDENTS SCORING AT EACH PERFORMANCE LEVEL

Standard (Kentucky/ NAEP)	Reading				Mathematics			
	Grade 4		Grade 8		Grade 4		Grade 8	
	Kentucky	NAEP	Kentucky	NAEP	Kentucky	NAEP	Kentucky	NAEP
Distinguished/ Advanced	0	2	0	n/a*	2	1	3	2
Proficient/ Proficient	3	19	8	n/a*	3	13	10	17
Apprentice/ Basic	47	55	71	n/a*	31	53	24	57
Novice/ Below Basic	50	45	21	n/a*	65	47	63	43

*Per Educational Testing Service policy, data from the National Assessment are not available at the state level in the manner necessary to complete this table.

CHAPTER 6
ITEM ANALYSIS

OVERVIEW

While content coverage is the predominant focus of advisory committees, one must also assure the technical quality of the assessment instruments. As indicated previously, analyses of field test results and results from the actual statewide administration of questions yield statistics of great use in the improvement and selection of test questions. Table 4-4 in Chapter 4 described the major item analyses performed. Tables 6-1 to 6-21 summarize the item analyses.

GRADE 4, OPEN-RESPONSE ITEMS

KIRIS open-response items are scored on a 0-4 scale, with 2.00 representing the middle of the difficulty range. Table 6-1 shows that in each year, all grade 4 tests are slightly difficult for the population of examinees, with most item difficulties falling between 1.21 and 2.00. The 1993-94 Mathematics and Science tests were more difficult than most, with a sixth and a third of the items, respectively, having difficulties below 1.21. However, Table 6-2 shows this greater difficulty did not negatively affect item-total correlations.

Table 6-3 shows that most items differentiated between *Novice* and non-*Novice* students reasonably well. On the other hand, Table 60 shows that many items did not differentiate well between *Apprentice* or below and *Proficient* and above. Table 61 shows that other than Mathematics in 1993-94, items did not differentiate well between *Distinguished* and below *Distinguished*. These results are appropriate. Given the current distribution of student proficiency levels, it is most critical that KIRIS differentiate well at the *Novice-Apprentice* boundary, somewhat important at the *Apprentice-Proficient* category, and not important at the *Proficient-Distinguished* boundary.

GRADE 8 OPEN-RESPONSE ITEMS

Table 6-8 shows that in each year, most grade 8 tests are slightly difficult for the population of examinees, with most item difficulties falling between 1.21 and 2.00. The 1993-94 Reading test and 1992-93 Social Studies test were exceptions, with a significant number of items with difficulties above. Table 6-9 shows these tests had marginally higher item-total correlations than most of the other tests.

Table 6-10 shows that most items differentiated between *Novice* and non-*Novice* students reasonably well. On the other hand, Table 6-11 shows that with the exception of Science, most items did a much better job differentiating between *Apprentice* or below and *Proficient* and above than did the grade 4 items. However, Table 6-12 shows that other than Mathematics in 1992-93 and 1993-94, items did not differentiate well between *Distinguished* and below *Distinguished*.

GRADE 12 OPEN-RESPONSE ITEMS

To facilitate comparisons across years, the 1991-92 item means in Table 6-15 are reported on a 0 to 4 scale rather than the 1 to 5 scale originally used. In each year most grade 12 tests were slightly difficult for the population of examinees, with most item difficulties falling between 1.21 and 2.00. The difficulty of the 1992-93 and 1993-94 Mathematics tests, was more extreme, with a large number of items (about 60 percent) with difficulties below .20. Table 6-16 shows these variations from medium difficulty did not negatively affect item-total correlations.

Table 6-17 shows that most items differentiated between *Novice* and non-*Novice* students reasonably well. On the other hand, Table 6-18 shows that similar to grade 8, with the exception of Science, most items did a much better job differentiating between *Apprentice* or below and *Proficient* and above than did the grade 4 items. However, Table 6-19 shows that other than Mathematics in 1992-93 and 1993-94, items did not differentiate well between *Distinguished* and below *Distinguished*.

TABLE 6-1
DISTRIBUTION OF ITEM SCORES FOR OPEN-RESPONSE ITEMS, GRADE 4

Subject Area	Year	Item Scores										N Items
		.00-.40	.41-.80	.81-1.20	1.21-1.60	1.61-2.00	2.01-2.40	2.41-2.80	2.81-3.20	3.21-3.60	3.61-4.00	
Mathematics	1991-92	0	0	1	6	7	1	0	0	0	0	15
	1992-93	0	0	2	12	14	1	0	0	0	0	29
	1993-94	0	0	5	12	10	2	0	0	0	0	29
Reading	1991-92	0	0	0	13	2	0	0	0	0	0	15
	1992-93	0	0	0	10	17	2	0	0	0	0	29
	1993-94	0	1	2	10	14	1	1	0	0	0	29
Science	1991-92	0	0	0	3	9	2	1	0	0	0	15
	1992-93	0	0	0	5	18	5	1	0	0	0	29
	1993-94	0	2	7	4	13	2	1	0	0	0	29
Social Studies	1991-92	0	0	0	2	9	3	1	0	0	0	15
	1992-93	0	0	0	5	16	7	1	0	0	0	29
	1993-94	0	0	0	10	11	8	0	0	0	0	29

TABLE 6-2
DISTRIBUTION OF ITEM-TOTAL CORRELATIONS FOR OPEN-RESPONSE ITEMS, GRADE 4

Subject Area	Year	Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	0	2	9	4	0	0	0	0	15
	1992-93	0	0	0	0	17	12	0	0	0	0	29
	1993-94	0	0	0	2	6	20	1	0	0	0	29
Reading	1991-92	0	0	0	1	3	11	0	0	0	0	15
	1992-93	0	0	0	0	3	26	0	0	0	0	29
	1993-94	0	0	0	0	7	21	1	0	0	0	29
Science	1991-92	0	0	0	4	9	2	0	0	0	0	15
	1899	0	0	0	6	22	1	0	0	0	0	29
	1993-94	0	0	0	4	14	11	0	0	0	0	29
Social Studies	1899	0	0	0	0	13	2	0	0	0	0	15
	1992-93	0	0	0	3	18	8	0	0	0	0	29
	1993-94	0	0	0	0	16	13	0	0	0	0	29

TABLE 6-3
NOVICE VS. APPRENTICE, PROFICIENT, DISTINGUISHED DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 4

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	0	7	5	1	2	0	0	0	15
	1992-93	0	0	0	12	13	3	1	0	0	0	29
	1993-94	0	0	0	7	15	4	3	0	0	0	29
Reading	1991-92	0	0	0	5	8	1	1	0	0	0	15
	1992-93	0	0	0	15	12	2	0	0	0	0	29
	1993-94	0	0	0	15	11	2	1	0	0	0	29
Science	1991-92	0	0	4	6	2	1	2	0	0	0	15
	1992-93	0	0	3	21	2	3	0	0	0	0	29
	1993-94	0	0	3	15	7	3	1	0	0	0	29
Social Studies	1991-92	0	0	2	9	1	2	1	0	0	0	15
	1992-93	0	0	2	18	8	0	1	0	0	0	29
	1993-94	0	0	1	14	10	4	0	0	0	0	29

TABLE 6-4
NOVICE, APPRENTICE VS. PROFICIENT, DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 4

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	2	10	1	2	0	0	0	0	0	15
	1992-93	0	2	17	8	2	0	0	0	0	0	29
	1993-94	0	2	5	17	5	0	0	0	0	0	29
Reading	1991-92	0	1	12	1	1	0	0	0	0	0	15
	1992-93	0	0	17	9	3	0	0	0	0	0	29
	1993-94	0	0	13	12	3	1	0	0	0	0	29
Science	1991-92	2	10	2	1	0	0	0	0	0	0	15
	1992-93	10	17	2	0	0	0	0	0	0	0	29
	1993-94	0	17	11	1	0	0	0	0	0	0	29
Social Studies	1991-92	0	3	8	2	2	0	0	0	0	0	15
	1992-93	0	2	22	2	3	0	0	0	0	0	29
	1993-94	0	1	11	12	5	0	0	0	0	0	29

TABLE 6-5
NOVICE, APPRENTICE, PROFICIENT VS. DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 4

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	9	5	1	0	0	0	0	0	0	15
	1992-93	0	13	14	2	0	0	0	0	0	0	29
	1993-94	0	5	20	4	0	0	0	0	0	0	29
Reading	1991-92	15	0	0	0	0	0	0	0	0	0	15
	1992-93	12	17	0	0	0	0	0	0	0	0	29
	1993-94	17	12	0	0	0	0	0	0	0	0	29
Science	1991-92	9	0	0	0	0	0	0	0	0	0	9
	1992-93	29	0	0	0	0	0	0	0	0	0	29
	1993-94	12	16	1	0	0	0	0	0	0	0	29
Social Studies	1991-92	9	6	0	0	0	0	0	0	0	0	15
	1992-93	29	0	0	0	0	0	0	0	0	0	29
	1993-94	11	18	0	0	0	0	0	0	0	0	29

TABLE 6-6
DISTRIBUTION OF PERCENT CORRECT FOR MULTIPLE-CHOICE ITEMS, GRADE 4

Subject Area	Year	Percent Correct										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	3	4	7	23	37	30	45	31	21	8	209
	1992-93	1	1	1	16	23	22	28	20	4	0	116
	1993-94	0	1	2	11	19	24	27	23	9	0	116
Reading	1991-92	0	0	11	29	61	64	29	6	1	0	201
	1992-93	0	0	3	7	22	32	38	11	3	0	116
	1993-94	0	0	1	4	19	27	39	22	4	0	116
Science	1991-92	1	6	18	16	38	50	41	34	14	2	220
	1992-93	1	7	10	10	22	24	21	16	5	0	116
	1993-94	0	1	13	18	24	27	13	13	7	0	116
Social Studies	1991-92	0	1	11	30	34	47	46	38	12	1	220
	1992-93	0	0	2	9	19	28	29	25	4	0	116
	1993-94	0	0	4	11	23	27	25	21	5	0	116

TABLE 6-7
DISTRIBUTION OF BISERIAL CORRELATIONS FOR MULTIPLE-CHOICE ITEMS, GRADE 4

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	8	21	61	79	36	4	0	0	0	0	209
	1992-93	1	5	36	57	15	1	0	0	0	0	115
	1993-94	1	9	23	58	25	0	0	0	0	0	116
Reading	1991-92	3	9	26	57	102	4	0	0	0	0	201
	1992-93	2	6	19	61	28	0	0	0	0	0	116
	1993-94	0	5	20	57	33	1	0	0	0	0	116
Science	1991-92	15	31	65	72	34	3	0	0	0	0	220
	1992-93	12	24	40	33	7	0	0	0	0	0	116
	1993-94	7	24	34	45	5	1	0	0	0	0	116
Social Studies	1991-92	5	31	55	89	40	0	0	0	0	0	220
	1992-93	0	10	29	51	26	0	0	0	0	0	116
	1993-94	4	14	31	42	24	1	0	0	0	0	116

TABLE 6-8
DISTRIBUTION OF ITEM SCORES FOR OPEN-RESPONSE ITEMS, GRADE 8

Subject Area	Year	Item Scores										N Items
		.00-.40	.41-.80	.81-1.20	1.21-1.60	1.61-2.00	2.01-2.40	2.41-2.80	2.81-3.20	3.21-3.60	3.61-4.00	
Mathematics	1991-92	0	0	1	6	5	0	2	1	0	0	15
	1992-93	0	0	4	12	11	0	2	0	0	0	29
	1993-94	2	5	8	7	7	0	0	0	0	0	29
Reading	1991-92	0	0	0	0	13	2	0	0	0	0	15
	1992-93	0	0	1	5	20	3	0	0	0	0	29
	1993-94	0	1	1	4	13	10	0	0	0	0	29
Science	1991-92	0	0	1	9	5	0	0	0	0	0	15
	1992-93	0	0	3	16	8	2	0	0	0	0	29
	1993-94	0	1	8	16	3	1	0	0	0	0	29
Social Studies	1991-92	0	0	1	8	6	0	0	0	0	0	15
	1992-93	0	0	0	2	21	6	0	0	0	0	29
	1993-94	1	0	7	14	6	1	0	0	0	0	29

TABLE 6-9
DISTRIBUTION OF ITEM-TOTAL CORRELATIONS FOR OPEN-RESPONSE ITEMS, GRADE 8

Subject Area	Year	Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	0	0	12	3	0	0	0	0	15
	1992-93	0	0	0	1	3	23	2	0	0	0	29
	1993-94	0	0	0	1	10	18	0	0	0	0	29
Reading	1991-92	0	0	0	0	1	9	5	0	0	0	15
	1992-93	0	0	0	0	2	11	16	0	0	0	29
	1993-94	0	0	0	0	3	7	19	0	0	0	29
Science	1991-92	0	0	0	2	9	4	0	0	0	0	15
	1992-93	0	0	0	0	13	16	0	0	0	0	29
	1993-94	0	0	0	6	19	4	0	0	0	0	29
Social Studies	1991-92	0	0	0	0	3	11	1	0	0	0	15
	1992-93	0	0	0	0	1	15	13	0	0	0	29
	1993-94	0	0	0	0	1	10	17	0	0	0	28

TABLE 6-10
NOVICE VS. APPRENTICE, PROFICIENT, DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 8

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	0	9	4	0	2	0	0	0	15
	1992-93	0	0	1	6	17	2	3	0	0	0	29
	1993-94	0	0	1	6	17	4	1	0	0	0	29
Reading	1991-92	0	0	0	5	7	3	0	0	0	0	15
	1992-93	0	0	0	14	14	1	0	0	0	0	29
	1993-94	0	0	2	9	14	4	0	0	0	0	29
Science	1991-92	0	0	2	8	2	1	2	0	0	0	15
	1992-93	0	0	1	13	12	1	2	0	0	0	29
	1993-94	0	0	2	17	6	3	1	0	0	0	29
Social Studies	1991-92	0	0	0	5	7	2	1	0	0	0	15
	1992-93	0	0	0	4	21	3	1	0	0	0	29
	1993-94	0	0	0	4	16	6	2	0	0	0	28

TABLE 6-11
NOVICE, APPRENTICE VS. PROFICIENT, DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 8

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	5	8	0	1	1	0	0	0	15
	1992-93	0	0	0	13	12	1	3	0	0	0	29
	1993-94	0	0	1	13	13	2	0	0	0	0	29
Reading	1991-92	0	0	6	7	2	0	0	0	0	0	15
	1992-93	0	0	3	14	11	1	0	0	0	0	29
	1993-94	0	0	1	12	14	2	0	0	0	0	29
Science	1991-92	5	9	1	0	0	0	0	0	0	0	15
	1992-93	0	15	11	3	0	0	0	0	0	0	29
	1993-94	2	21	6	0	0	0	0	0	0	0	29
Social Studies	1991-92	0	0	2	10	2	1	0	0	0	0	15
	1992-93	0	0	2	15	10	2	0	0	0	0	29
	1993-94	0	0	1	8	16	3	0	0	0	0	28

TABLE 6-12
NOVICE, APPRENTICE, PROFICIENT VS. DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 8

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	12	1	2	0	0	0	0	0	0	15
	1992-93	0	0	14	10	4	1	0	0	0	0	29
	1993-94	0	1	11	16	1	0	0	0	0	0	29
Reading	1991-92	11	4	0	0	0	0	0	0	0	0	15
	1992-93	14	13	2	0	0	0	0	0	0	0	29
	1993-94	1	27	1	0	0	0	0	0	0	0	29
Science	1991-92	7	0	0	0	0	0	0	0	0	0	7
	1992-93	27	2	0	0	0	0	0	0	0	0	29
	1993-94	19*	0	0	0	0	0	0	0	0	0	29
Social Studies	1991-92	3	12	0	0	0	0	0	0	0	0	15
	1992-93	1	21	7	0	0	0	0	0	0	0	29
	1993-94	0	14	14	0	0	0	0	0	0	0	28

*Because only 23 students out of 51621 were classified as Distinguished in Science, several of the 12 forms had no Distinguished students. No biserial correlations could be computed for either of the two items on each of five forms, and therefore correlations are reported for only 19 of 29 items.

TABLE 6-13
DISTRIBUTION OF PERCENT CORRECT FOR MULTIPLE-CHOICE ITEMS, GRADE 8

Subject Area	Year	Percent Correct										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	9	10	36	79	55	18	10	2	0	219
	1992-93	0	4	5	24	27	28	18	8	2	0	116
	1993-94	0	2	7	28	22	22	19	10	0	0	116
Reading	1991-92	0	2	7	27	35	75	52	35	8	0	241
	1992-93	1	1	1	11	20	28	25	23	6	0	116
	1993-94	0	1	2	10	13	24	34	24	8	0	116
Science	1991-92	0	0	11	19	35	37	24	8	2	0	136
	1992-93	0	1	19	30	28	20	14	4	0	0	116
	1993-94	0	2	19	29	28	21	15	2	0	0	116
Social Studies	1991-92	5	11	9	27	59	72	34	36	3	0	256
	1992-93	0	1	12	23	22	25	16	13	4	0	116
	1993-94	0	0	12	30	21	22	12	15	4	0	116

TABLE 6-14 DISTRIBUTION OF BISERIAL CORRELATIONS FOR MULTIPLE-CHOICE ITEMS, GRADE 8												
Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	6	30	80	64	36	3	0	0	0	0	219
	1992-93	3	8	37	58	10	0	0	0	0	0	116
	1993-94	1	4	25	66	20	0	0	0	0	0	116
Reading	1991-92	5	13	60	100	52	11	0	0	0	0	241
	1992-93	3	12	33	53	14	1	0	0	0	0	116
	1993-94	1	9	20	41	43	2	0	0	0	0	116
Science	1991-92	6	22	47	50	11	0	0	0	0	0	136
	1992-93	19	36	42	19	0	0	0	0	0	0	116
	1993-94	15	29	47	24	1	0	0	0	0	0	116
Social Studies	1991-92	24	27	61	86	53	5	0	0	0	0	256
	1992-93	10	17	30	48	11	0	0	0	0	0	116
	1993-94	7	13	31	34	30	1	0	0	0	0	116

TABLE 6-15
DISTRIBUTION OF ITEM SCORES FOR OPEN-RESPONSE ITEMS, GRADE 12

Subject Area	Year	Item Scores										N Items
		.00-.40	.41-.80	.81-1.20	1.21-1.60	1.61-2.00	2.01-2.40	2.41-2.80	2.81-3.20	3.21-3.60	3.61-4.00	
Mathematics	1991-92	0	0	7	6	1	1	0	0	0	0	15
	1992-93	0	1	15	9	4	0	0	0	0	0	29
	1993-94	2	7	9	8	3	0	0	0	0	0	29
Reading	1991-92	0	0	1	10	4	0	0	0	0	0	15
	1992-93	0	0	1	11	13	4	0	0	0	0	29
	1993-94	0	1	0	6	14	8	0	0	0	0	29
Science	1991-92	0	0	0	5	9	1	0	0	0	0	15
	1992-93	0	0	4	22	3	0	0	0	0	0	29
	1993-94	0	1	8	12	7	1	0	0	0	0	29
Social Studies	1991-92	0	0	1	10	2	2	0	0	0	0	15
	1992-93	0	0	6	16	6	1	0	0	0	0	29
	1993-94	0	2	0	10	13	4	0	0	0	0	29

TABLE 6-16
DISTRIBUTION OF ITEM-TOTAL CORRELATIONS FOR OPEN-RESPONSE ITEMS, GRADE 12

Subject Area	Year	Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	0	1	10	4	0	0	0	0	15
	1992-93	0	0	0	1	4	14	10	0	0	0	29
	1993-94	0	0	0	2	6	17	4	0	0	0	29
Reading	1991-92	0	0	0	0	0	3	12	0	0	0	15
	1992-93	0	0	0	0	2	13	14	0	0	0	29
	1993-94	0	0	0	2	3	14	10	0	0	0	29
Science	1991-92	0	0	0	2	5	4	4	0	0	0	15
	1992-93	0	0	0	0	6	19	4	0	0	0	29
	1993-94	0	0	1	0	9	14	5	0	0	0	29
Social Studies	1991-92	0	0	0	0	1	5	9	0	0	0	15
	1992-93	0	0	0	0	2	9	18	0	0	0	29
	1993-94	0	0	0	0	1	8	20	0	0	0	29

TABLE 6-17
NOVICE VS. APPRENTICE, PROFICIENT, DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 12

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	4	5	4	1	0	0	1	0	15
	1992-93	0	0	1	12	12	4	0	0	0	0	29
	1993-94	0	0	5	14	8	0	2	0	0	0	29
Reading	1991-92	0	0	0	0	7	6	2	0	0	0	15
	1992-93	0	0	0	3	19	7	0	0	0	0	29
	1993-94	0	0	1	5	17	6	0	0	0	0	29
Science	1991-92	0	0	2	6	4	1	2	0	0	0	15
	1992-93	0	0	0	13	12	4	0	0	0	0	29
	1993-94	0	0	2	8	17	1	1	0	0	0	29
Social Studies	1991-92	0	0	0	2	9	1	3	0	0	0	15
	1992-93	0	0	0	4	15	7	3	0	0	0	29
	1993-94	0	0	1	3	13	11	1	0	0	0	29

TABLE 6-18
NOVICE, APPRENTICE VS. PROFICIENT, DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 12

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	0	3	7	4	0	1	0	0	0	15
	1992-93	0	1	3	3	16	4	2	0	0	0	29
	1993-94	0	0	3	6	13	6	0	1	0	0	29
Reading	1991-92	0	0	0	9	3	3	0	0	0	0	15
	1992-93	0	2	12	14	1	0	0	0	0	0	29
	1993-94	0	0	1	19	6	3	0	0	0	0	29
Science	1991-92	0	4	6	4	1	0	0	0	0	0	15
	1992-93	0	7	18	4	0	0	0	0	0	0	29
	1899	0	2	13	13	1	0	0	0	0	0	29
Social Studies	1991-92	0	1	3	6	4	1	0	0	0	0	15
	1992-93	0	1	15	10	3	0	0	0	0	0	29
	1993-94	0	0	0	11	13	5	0	0	0	0	29

TABLE 6-19
NOVICE, APPRENTICE, PROFICIENT VS. DISTINGUISHED
DISTRIBUTION OF BISERIAL CORRELATIONS FOR OPEN RESPONSE ITEMS, GRADE 12

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	0	7	6	1	1	0	0	0	0	0	15
	1992-93	0	1	10	12	3	3	0	0	0	0	29
	1993-94	0	1	7	15	5	0	1	0	0	0	29
Reading	1991-92	1	12	2	0	0	0	0	0	0	0	15
	1992-93	6	21	2	0	0	0	0	0	0	0	29
	1993-94	0	18	10	1	0	0	0	0	0	0	29
Science	1991-92	8	7	0	0	0	0	0	0	0	0	15
	1992-93	13	14	0	0	0	0	0	0	0	0	27
	1993-94	5	22	2	0	0	0	0	0	0	0	29
Social Studies	1991-92	6	9	0	0	0	0	0	0	0	0	15
	1992-93	11	16	2	0	0	0	0	0	0	0	29
	1993-94	0	16	12	1	0	0	0	0	0	0	29

TABLE 6-20
DISTRIBUTION OF PERCENT CORRECT FOR MULTIPLE-CHOICE ITEMS, GRADE 12

Subject Area	Year	Percent Correct										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	1	13	30	38	57	40	25	11	4	0	219
	1992-93	1	3	15	28	27	21	15	4	2	0	116
	1993-94	2	6	19	23	22	19	16	5	4	0	116
Reading	1991-92	0	0	6	23	44	34	56	31	15	0	209
	1992-93	0	2	0	4	14	17	26	30	22	0	115
	1993-94	0	0	1	9	15	17	21	31	16	6	116
Science	1991-92	0	13	19	36	48	48	27	24	3	0	218
	1992-93	0	3	11	14	23	29	19	14	3	0	116
	1993-94	0	4	12	16	23	27	15	15	4	0	116
Social Studies	1991-92	2	8	17	55	48	47	37	29	13	0	256
	1992-93	1	2	3	14	19	25	29	13	10	0	116
	1993-94	1	2	6	10	26	21	24	17	9	0	116

TABLE 6-21
DISTRIBUTION OF BISERIAL CORRELATIONS FOR MULTIPLE-CHOICE ITEMS, GRADE 12

Subject Area	Year	Biserial Correlation										N Items
		.00-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-1.00	
Mathematics	1991-92	12	28	28	70	60	17	2	2	0	0	219
	1992-93	11	15	28	43	19	0	0	0	0	0	116
	1993-94	10	16	35	49	6	0	0	0	0	0	116
Reading	1991-92	2	2	22	47	69	44	16	6	1	0	209
	1992-93	4	8	26	50	27	0	0	0	0	0	115
	1993-94	1	9	33	52	11	0	0	0	0	0	116
Science	1991-92	14	18	48	72	50	10	5	1	0	0	218
	1992-93	9	14	37	40	16	0	0	0	0	0	116
	1993-94	13	10	38	44	11	0	0	0	0	0	116
Social Studies	1991-92	10	17	30	80	91	13	7	6	1	0	255
	1992-93	2	12	30	47	25	0	0	0	0	0	116
	1993-94	5	15	28	52	16	0	0	0	0	0	116

SECTION 3
ADMINISTRATION, SCORING, AND REPORTING

Section 3 deals with the operational heart of the KIRIS testing and accountability programs. Chapter 7 describes how all portions of the test were administered. Chapter 8 describes the scoring process for the test. Chapter 9 presents information about the equating process used to maintain standards across years and across test forms within years. Chapter 10 highlights the score reporting process – which reports were sent to schools and to districts and when were they sent. Chapter 11 presents the results for the first accountability cycle. Because of their complexity and controversy, scoring issues related to portfolios get a separate chapter of their own, Chapter 12.

CHAPTER 7
TEST ADMINISTRATION

DETERMINING STUDENTS FOR WHOM A SCHOOL IS ACCOUNTABLE

For the 1993-94 school year, schools were held accountable for open-response test scores of all students enrolled in the school on the first day of the on-demand testing window (Feb. 7, 1994, for grade 12; April 18, 1994, for grades 4 and 8). Schools were held accountable for a student's writing portfolio score if the student was enrolled in the school on the day writing portfolios were due, as long as that student had been enrolled in any Kentucky public school for at least 100 days. Previously, schools were held accountable for all students enrolled on the 20th day of school for 1992-93, or December 15,1991, for 1991-92. The need to balance issues of perceived fairness (i.e., holding schools accountable for students who moved soon after the 20th school day) with logistical difficulties associated with tracking students back to their original school prompted the adoption of the rules. Table 7-1 compares the accountability rules for each year of the first accountability cycle. Eligibility rules are not included for the assessments that were administered in 1992-93 and 1993-94 but did not count in the first accountability cycle, i.e., mathematics and alternate portfolios, and on-demand writing. For performance events, schools were held accountable for students attending school on the day of testing; this policy did not change across the years.

TABLE 7-1
RULES FOR DETERMINING STUDENTS FOR WHOM A SCHOOL IS ACCOUNTABLE

Accountability Year	On-Demand Eligibility	Writing Portfolio Eligibility
1991-92	Enrolled on December 15, 1991	Same as on-demand
1992-93	Enrolled on 20th day of school year	Same as on-demand
1993-94	Grades 4 and 8 – enrolled on April 18, 1994 Grade 12 – enrolled on February 7, 1994	Enrolled on March 1, 1994, and enrolled in one or more Kentucky public schools for at least 100 days by March 1

Any student for whom a school was accountable, but who was not tested or did not produce a portfolio, was assigned a score of *Novice*.

COLLECTING ENROLLMENT INFORMATION. For 1993-94, no advance collections of enrollment information was necessary. Student names printed on the demand test forms indicated day of testing enrollment. For 1991-92 and 1992-93, the enrollment information was collected in two ways. The majority of schools submitted enrollment information on scannable registration forms where students (or teachers) provided name, sex, race, date of birth, and information on the amount of time a language other than English was spoken in the student's home. School personnel completed sections that indicated if the student was eligible for Bilingual or English as a Second Language education or if the student had an identified educational disability. As a second option, some schools chose to submit student registration information on computer diskettes.

After receiving school rosters, Advanced Systems assigned a unique identification number to each student. This identification number was pre-printed on students' on-demand answer folders and portfolio score forms, and was used to match students' scores across the on-demand and portfolio components of the assessment program.

When schools enrolled new students during the course of the year, these students were added to the school files by completing blank (not pre-printed) student answer sheets at the time of on-demand testing and/or by submitting portfolio scores for new students. These students were appended to the file so that the school would receive results for them.

MODIFICATIONS TO DATA FILES. In order to verify 1993-94 student membership with regard to school accountability, rosters of students who were included in the September 1994 school reports were sent to every school prior to or concurrently with the release of the reports. Schools were instructed to note any changes that should be made to their rosters for accountability purposes; the superintendent and district assessment coordinator certified that any noted changes were approved. Student membership in previous years was verified in a similar procedure. The acceptable changes to the rosters involved circumstances such as the following:

- a student enrolled after the 20th day of instruction for 1992-93, or after December 15, 1991, for 1991-92: the school indicated which school the student previously attended. The student's open-response and portfolio results were reassigned to the accountable school — the one where the child was enrolled on the 20th day and/or December 15. No reassignment of performance events scores was made because, for those tests, schools were accountable for students enrolled on the day of testing.
- a student eligible for language exemption according to Department of Education policy: students whose native language was not English and who had been enrolled in an English-speaking school for fewer than two years were exempt from testing.
- a foreign exchange student.
- a student with an identified disability who was eligible for an alternate portfolio according to criteria established by the Department of Education: students in this category were exempt from on-demand testing, performance events, and writing and mathematics portfolios. These students each developed an alternate portfolio that will count towards second accountability cycle baselines. Criteria for eligible students were detailed in the Program Advisory 92-OCAA-104 (February, 1993) and published in the *Kentucky Alternate Portfolio Teacher's Guide* and the *1992-93 District Assessment Coordinator Implementation Guidebook*.
- a student from a non-accountable grade: in a number of cases, a school's student registration information was incorrect, showing students enrolled in grades other than 4, 8, or 12. Schools corrected any of these errors on their rosters.
- a student medically excluded from testing: a student could be excluded from testing for medical reasons if a signed doctor's statement was provided.

As certified changes to the rosters were received, new information was verified by the District Assessment Coordinators, (DAC's) and some cases were further investigated by Advanced Systems' and the Department of Education's project management staff. Confirmed changes were then made to the student master files by entering the appropriate code that would remove the student from calculations of a school's results for 1993-94, or by moving the student's record to the school where he or she was enrolled on December 15, 1991 for the 1991-92 school year, or on the 20th day of instruction for the 1992-93 school year.

Students who were not exempt but did not attempt the on-demand test or portfolio were assigned to the *Novice* performance level in the accountability reports. In 1993-94, about 0.2 percent of Kentucky's fourth graders, 0.6 percent of the eighth graders, and 0.4 percent of the twelfth grade students were described as "Novice" because they did not complete the tests. Tables 7-2 through 7-4 show for grades 4, 8, and 12, respectively, the number of students classified as exempt from on-demand testing, the number classified as Novice because they did not attempt the tests, and the number who completed the on-demand tests.

TABLE 7-2
GRADE 4 STUDENTS IN EACH ACCOUNTABILITY ELIGIBILITY CATEGORY

Accountability Eligibility Category	1991-92		1992-93		1993-94	
	Number	Percent	Number	Percent	Number	Percent
Exempted: Language	53	0.1	45	0.1	69	0.1
Exempted: Special Education	416	0.8	176	0.3	136	0.3
Exempted: Medical	0	0.0	113	0.2	35	0.1
Exempted: Foreign Exchange	4	0.0	19	0.0	4	0.0
Exempted: Other	1977	3.7	1868	3.6	270	0.5
Exempted: Total	2450	4.6	2221	4.3	510	1.0
Eligible: Not Tested (*Novice*)	2083	3.9	464	0.9	82	0.2
Eligible: Tested	48444	91.4	49559	94.9	49704	98.8
Eligible: Total	50527	95.4	50021	95.7	49786	99.0
Total	52977	100.0	52242	100.0	50296	100.0

TABLE 7-3
GRADE 8 STUDENTS IN EACH ACCOUNTABILITY ELIGIBILITY CATEGORY

Accountability Eligibility Category	1991-92		1992-93		1993-94	
	Number	Percent	Number	Percent	Number	Percent
Exempted: Language	15	0	23	0	24	0.0
Exempted: Special Education	252	0.5	246	0.5	157	0.3
Exempted: Medical	0	0.0	219	0.4	50	0.1
Exempted: Foreign Exchange	5	0.0	40	0.1	4	0.0
Exempted: Other	1398	2.7	1667	3.1	420	0.8
Exempted: Total	2109	4.0	2195	4.1	655	1.3
Eligible: Not Tested (*Novice*)	2873	5.5	1313	2.5	320	0.6
Eligible: Tested	47596	90.5	49537	93.4	51621	98.1
Eligible: Total	50469	96.0	50850	95.9	51941	98.8
Total	52578	100.0	53045	100.0	52596	100.0

TABLE 7-4
GRADE 12 STUDENTS IN EACH ACCOUNTABILITY ELIGIBILITY CATEGORY

Accountability Eligibility Category	1991-92		1992-93		1993-94	
	Number	Percent	Number	Percent	Number	Percent
Exempted: Language	12	0.0	12	0.0	14	0.0
Exempted: Special Education	109	0.3	86	0.2	116	0.3
Exempted: Medical	0	0.0	327	0.6	78	0.2
Exempted: Foreign Exchange	113	0.3	66	0.1	179	0.5
Exempted: Other	1012	2.7	1199	2.9	242	0.6
Exempted: Total	1246	3.4	1690	4.3	629	1.7
Eligible: Not Tested (*Novice*)	514	1.4	1827	4.6	164	0.4
Eligible: Tested	35332	95.3	36224	91.1	36597	97.9
Eligible: Total	35846	96.6	38051	95.7	36761	98.3
Total	37092	100.0	39751	100.0	37390	100.0

The exempted 'other' category includes the following cases where student results were not used in the accountability system:

- students who moved out of state (or to private schools) and were unavailable for testing;
- students who moved into Kentucky from an out-of-state school after the 20th day of instruction in the 1991-92 testing, or after December 15, 1991 in the 1991-92 testing;
- students who graduated early.

Due to differences in record keeping across years, a small number of additional students are included in the 'other' category in order to simplify the tables. Eligible students are noted above as 'not tested (novice)' in cases where no attempt has been made at any open-response question, or all open-response questions were given off-topic responses and no multiple-choice questions were attempted.

Final rosters listing the names and scores of students included in accountability calculations accompanied the schools' accountability reports.

ADMINISTRATION OF OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS

GENERAL PROCEDURES. Table 7-5 presents the test dates for each year of the first accountability cycle.

TABLE 7-5
TESTING WINDOWS*

Grade	Test Type	Year		
		1991-92	1992-93	1993-94
4	On-demand	April 13 - May 1	April 19 - 30	April 19 - 30
	Performance Events	April 15 - May 8	April 7 - 23	April 11 - 29
8	On-demand	April 13 - May 1	April 19 - 30	April 19 - 30
	Performance Events	May 4 - May 15	April 19 - May 7	April 25 - May 13
12	On-demand	February 10 - 28	February 8 - 19	February 8 - 19
	Performance Events	March 23 - 27	March 1 - 12	March 7 - 18

*Not including make-up testing

As an example, for the 1993-94 school year, testing of all grade 12 students in Kentucky public schools occurred between February 8 and 19, 1994. Testing for grades 4 and 8 took place April 19 through 30. Within these testing windows, schools were allowed to set up their own specific testing schedules. Coordinators' and administrators' manuals served as a guide to administration. The manuals detailed timing requirements, directions for students, and other considerations for administering the tests and handling materials. *Appropriate Assessment Practices,* a document outlining appropriate behaviors for school personnel during testing, was included in the test administration manual; a copy of this document is available from the Kentucky Department of Education.

Four subject areas were assessed using open-response and multiple-choice tests: reading, mathematics, science, and social studies. Arts and humanities, practical living, and vocational studies were integrated into some open-response items and writing prompts. (Note that data from the arts and humanities, practical living, and vocational studies items are not included in figures for this accountability cycle.) In 1992-93, each subject was administered in one 90-minute test session at grades 8 and 12; students were given up to 45 minutes' additional time, if needed, to complete the test. The grade 4 test sessions were 60 minutes long, and students were given up to 30 additional minutes to complete a session. The additional time was scheduled directly after the initial test session for the subject area. Test administration procedures in 1993-94 provide students with as much additional time as needed as long as constructive progress is being made toward completion of the test.

Students received a test booklet and a separate scannable answer document in which they recorded their answers to all questions, including open-response questions. A sample answer document is included in Appendix 7-1. Each student's name and identification number were pre-printed on the cover of his or her answer document; additional blank folders (without pre-printed information) were provided for students who entered school after enrollments were collected by Advanced Systems. Twelve forms of the test were created; students were randomly given one form that they used for all sections of the test. In other words, if they received Form 1 for the reading test, they completed Form 1 for other subject areas as well.

Each school principal completed a *Principal's Certification of Proper Test Administration* to confirm enrollment and testing figures. The principal also certified on this form that the testing was conducted in accordance with the instructions provided in the manuals and that all materials were handled in a secure manner. A copy of the form is presented in Appendix 7-2, along with a similar form completed by District Assessment Coordinators.

SHIPPING AND RECEIVING PROCEDURES. All materials were packaged by the school and shipped to District Assessment Coordinators. The materials were sent to district offices approximately one week prior to the beginning of test administration. All district and school personnel were informed that the materials were secure and that all materials had to be returned to Advanced Systems at the completion of testing.

District Assessment Coordinators were instructed to package all materials from all schools in their districts within one week after the completion of testing. The shipping company picked up packaged materials from every district office.

As packages of answer sheets were received at Advanced Systems, they were collected in a prescribed location for check-in. Packages were opened one at a time, answer sheets were counted, and the count was compared with information provided by the school principal on the *Principal's Certification of Proper Test Administration*. The count was entered in a log, and any discrepancies were noted both on the log and on the principal's form.

To prepare the answer sheets for machine- and hand-scoring, the cover pages, containing multiple-choice responses, were separated from the open-response pages, and sorted by form. The two different types of responses were boxed separately. Finally, within each box of open-response forms, answer sheets were shuffled to ensure that multiple scorers would evaluate open responses from each school.

Test booklets and other material were checked for stray answer documents or other forms mistakenly shipped with them. The test booklets were then destroyed.

ADMINISTRATION OF PERFORMANCE EVENTS

As in previous years, 1993-94 performance events testing was managed by Advanced Systems' office in Louisville, Kentucky. In addition to the Regional Director and the Supervisor of Project Support Services, the staff included four regional events managers and a cadre of 73 specially trained facilitators who traveled to each school to administer the event tasks.

Tentative dates were sent to school principals by mid-October, requests for changes were accepted and accommodated, and final test date confirmation were mailed before Thanksgiving. Principals were asked to designate a school contact who would serve as ASME's point of communication for all future interchange with regard to performance events.

After scheduling was completed, events managers were responsible for coordinating all facets of the events administration in a specific geographic area of the state. Their duties included:

- individual phone consultations with designated contacts at each school to confirm schedules and a plan individual school testing points

- checking the list of schools included for testing for correctness and completeness;
- recruiting, interviewing and hiring test facilitators (each manager hired 17-18 facilitators from his/her region, to cover schools within that region);
- developing facilitator work schedules;
- monitoring and trouble-shooting throughout the weeks of events administration (examples: replacing or substituting for facilitators who were unable to meet prescribed schedules; re-scheduling test dates cancelled due to unforeseen circumstances; responding to school/district calls for information/clarification, record-keeping, etc).

Additional temporary personnel were employed by the Kentucky office during the events administration to fill roles including clerical support and light labor required to pack performance events testing materials.

Applicants for the facilitator positions were recruited through the following mechanisms:

- position announcements were mailed to approximately 106 local chapters of the Kentucky Retired Teachers Association for distribution to its membership;
- notices sent to placement offices of Kentucky State Colleges and Universities
- the Kentucky Department of Employment Services listed the position in its statewide job bank and authorized Advanced Systems' staff to conduct interviews at their 14 sites across the state;
- advertisements were placed in the Courier-Journal (state edition); the Owensboro Messenger-Inquirer; the Lexington Herald-Leader; the Paducah Sun-Democrat; Ohio Country Messengers
- notices sent to the urban League
- individual contacts with school boards, public school educators, colleges and universities, and community groups.

Managers screened hundreds of applications and conducted approximately 250 interviews to hire 73 facilitators. Hiring decisions were based on applicant qualifications in accordance with the requirements set forth in the position announcement; preference was given to applicants with bachelor's degrees and/or experience in school classrooms.

Table 7-6 profiles the 73 facilitators employed to administer the Performance Events.

TABLE 7-6
PROFILE OF PERFORMANCE EVENTS FACILITATOR QUALIFICATIONS AND DEMOGRAPHICS

Education	Degrees beyond the Baccalaureate	34
	Bachelor's Degree	36
	Associate's Degree	2
	Some college study	1
Experience	School principal	3
	Teacher	27
	School aide	22
	Other/non-school	21
Demographics	Male	23
	Female	50
	Black	2
	White	71

TRAINING OF FACILITATORS. Training for facilitators was conducted by Advanced Systems' Kentucky staff, with assistance from the KIRIS Assistant Project Director from the New Hampshire office. Staff members from the Kentucky Department of Education also attended. Training was conducted in two intensive two-day sessions in Louisville.

During the first of the training sessions, facilitators were provided with an overview of the Kentucky Education Reform Act; an introduction to the philosophy and purposes of performance assessment; extensive practice with the different assessment tasks, kits, and administration directions for the grade level; briefings on school protocol; and management of logistical issues such as travel, packing, and collection and return of student response forms. Concerns which could affect test administration and its place in the accountability measure were also examined and discussed. These issues included: the testing of special needs students; random assignment of students to groups; the role of observers/teachers in the administration; and appropriate responses to difficult, unpredictable or attitudinal problems in the schools. At the conclusion of the first three days of training, facilitators received testing kits and student response forms for grade 12 administration, and returned to their homes for three days of individual study and preparation before beginning testing.

Facilitators returned to Louisville after the conclusion of grade 12 testing for two more days of training in preparation for grades 4 and 8 events. During this training session, facilitators discussed their experiences with grade 12 and shared information that could be applied to testing at the lower grades. Administration guidelines for grades 4 and 8 were similar to those for grade 12 (i.e., random groupings, inclusion of all students, security of tasks, etc.) but there were unique modifications for each grade level that were explored during the additional training. As with grade 12, facilitators had extensive hands-on practice with the 27 new tasks — 13 at grades 4 and 14 at Grade 8. Again, at the conclusion of training, facilitators received their testing kits and student response forms, and returned to their homes for a few days of individual study and preparation before beginning testing.

ADMINISTRATION PROCEDURES. Advanced Systems' facilitators delivered materials necessary for performance test administration, including Student Response Forms and the unique supplies and manipulatives required for each task, on the scheduled day of testing for each school. Facilitators worked with designated school contacts to conduct testing following operational directions and policy. Facilitators set up materials for each testing session, randomly assigned students in each session to a task group, gave general directions and timed the tasks, collected all testing materials, and prepared "Session Reports" and "Participation Rate" forms. At the conclusion of testing, facilitators collected all student products and returned them to the New Hampshire office of Advanced Systems for scoring.

Performance events were administered to all students in attendance on the day of testing in grades 4, 8, and 12 at each school designated "A-1", a classification indicating full accountability as a Kentucky school. The events at each grade level included math, science, social studies, arts and humanities, practical living, and vocational studies. Performance events are summarized in Appendix 7-3. The tasks were typically designed for groups of three to four students. In keeping with KERA, schools were to have virtually all enrolled students participating in the events; specific exclusions were detailed in the "Commissioner's Program Advisories" of January 31, 1992 and February 9, 1993.

In general, students participated in only one event and in only one testing session. To minimize the sharing of information between students tested early in the day and those tested later, the testing at grades 8 and 12 was arranged so that all students who completed a particular event did so at the same time sharing of information Grade 4 was not likely to be considered a problem due to students being tested with their self-contained class, so the administration was arranged so that half of the events were taken by each classroom. In the case of very small schools (i.e., fewer than approximately 36 students at the grade level), students participated in two events during two sessions so that as many different events as possible could be administered in the school.

Schools were asked to provide a room or rooms in which the testing could be conducted, a teacher or other member of the instructional staff to help maintain order and observe the process, and specific standard materials (equipment and resources typically found in schools) needed for the tasks. Principals had the option of allowing additional members of their instructional staff to observe each testing session (two was the recommended maximum number). Principals were asked to ensure that students who used special equipment in their regular instructional programs (as specified in an IEP or 504 plan) had those resources available during testing, and to identify children with special physical needs so that facilitators could make appropriate task assignments.

MAKE-UP SESSIONS. Make-up sessions for an entire school were held only if unforeseen circumstances (e.g., school closings due to inclement weather) prevented testing on the originally scheduled date. There was no make-up testing for students who were not present on the test day. Schools were held accountable for the performance of students present in school on the day of testing, and were expected to have a participation rate of at least 90% of the students enrolled in the school on the day of testing. Using enrollment data supplied by principals, facilitators calculated participation rates before leaving the school at the end of each testing day, leaving that information with principals. KDE was sent a copy of the participation rate forms for those schools with less than 90% participation; these schools were required to provide an explaination to the Department of Education.

ADMINISTRATION OF WRITING AND MATHEMATICS PORTFOLIOS

KIRIS writing portfolios were administered in much the same way in 1993-94 as in previous years. A detailed description of the structure and process of training for development and scoring portfolios can be found in the *1991-92 Technical Report* and in *The Writing Portfolio Development and Scoring Process*, distributed in March, 1993. In 1992-93, mathematics portfolios were added to the KIRIS assessment in the accountability grades. The entire mathematics portfolio process followed the structure and procedures established for writing portfolios. There were no major changes from 1992-93 to 1993-94.

TEACHER TRAINING. The training model for 1992-93 portfolio development and scoring followed the pyramidal region/cluster organization established for 1991-92. The state was divided into nine regions, with each region assigned a Regional Coordinator for each accountability grade. Regional Coordinators participated in training Cluster Leaders in their regions and served as members of their respective advisory committees. Cluster Leaders are the teachers who trained all participating teachers, in "clusters" of 20–25, to develop and score portfolios. For writing portfolios, District Assessment Coordinators (DACs) in local districts selected the Cluster Leaders. For mathematics portfolios, Cluster Leaders made application to and were selected by the Department of Education.

Most initial training focused on developing and implementing portfolios in the classroom. Later training tended to focus on scoring portfolios.

TRAINING FOR DEVELOPMENT AND IMPLEMENTATION. The first phase of training addressed the development of portfolios, fulfillment of requirements, and state guidelines for the generation of student work. For both writing and mathematics, Regional Coordinators were the first to receive this intensive training. For writing portfolios, Regional Coordinators then provided the same training for Cluster Leaders. Advanced Systems and Kentucky Department of Education consultants, assisted by Regional Coordinators, trained the Cluster Leaders. Cluster Leaders then returned to their local districts to provide the same training to all classroom teachers involved in the development of either type of portfolio. For mathematics portfolio development, KDE and Advanced Systems' mathematics consultants also conducted training that was telecast by Kentucky Educational Television (KET). Districts were encouraged to record the telecasts, or could order video cassette directly from KET. The following materials were used during the first phase of training.

Writing Portfolios:

- *The Kentucky Writing Portfolio Teacher Handbook*
- "The Kentucky Writing Portfolio Holistic Scoring Guide"
- "Assessment Update: Portfolio Panel," aired October 7, 1992 (videocassette available)

Mathematics Portfolios:

- *The Kentucky Mathematics Portfolio Teacher's Guide*
- Sample Portfolios
- *Kentucky's Preliminary Curriculum Framework* (Goal II, pp. 179-208)
- "Assessment Update 5: Getting Ready for Math Portfolios," aired May 6, 1992 (videocassette available)
- "Assessment Update: Getting Started with Math Portfolios" (cluster training), aired September 2, 1992 (videocassette available)
- "Assessment Update: Portfolio Panel," aired October 7, 1992 (videocassette available)

TRAINING FOR SCORING. Teachers were trained on the standards and procedures for scoring writing and/or mathematics portfolios. This was accomplished through the same training procedures as for development and implementation. Telecasts were also used for this part of mathematics portfolio training. The following materials were used during this second phase of training.

Writing Portfolios:

- *The Kentucky Writing Portfolio Teacher Handbook*
- "The Kentucky Writing Portfolio Holistic Scoring Guide"
- *The 1993-94 Kentucky Writing Portfolio Scoring Training Manual*

Mathematics Portfolios:

- *The Kentucky Mathematics Portfolio Teacher's Guide*
- "The Kentucky Mathematics Portfolio Holistic Scoring Guide"
- *KIRIS Mathematics Portfolio Assessment Scoring Training Manual*
- "Assessment Update: Timeline for Scoring Math Portfolios" (cluster training), aired March 3, 1993 (videocassette available)
- "Mathematics Portfolios Scoring Training," Overview and Grades 4, 8, and 12, aired June 16, 1993 (four separate videocassettes available)

ADDITIONAL TRAINING AND MATERIALS. In addition to the two established phases of training for all participating teachers in the state, additional optional training and supplemental materials, described below, were offered throughout the 1992-93 school year. In all regions of the state, KDE and Advanced Systems conducted optional guided scoring sessions to provide a supportive scoring environment and a forum for discussion of portfolio issues and scoring standards for both mathematics and writing portfolios. For writing, these sessions took place during the first week of the scoring window in early March, 1993. For mathematics, the sessions were held during the first two weeks of August, 1993.

With regard to Writing Portfolios, a question-and-answer brochure, addressing some of the most commonly-asked questions from teachers, was compiled by the Department of Education and Advanced Systems consultants and distributed to writing portfolio cluster leaders. Also, the Writing Advisory Committee met to score a stratified random sample of portfolios and identify a set of Consensus Portfolios reflecting state standards. These were used to train and qualify scorers who participated in an important series of portfolio rescoring sessions described in Chapter 12.

With regard to Mathematics Portfolios, all teachers of mathematics at an accountability teachers were provided with addenda to the *Kentucky Mathematics Portfolio Teacher's Guide* that consisted of "Guidelines for the Generation of Student Work," two additional sample portfolio tasks per accountability grade, "Self-Assessment Conference Form," a bibliography, and a question-and-answer document that addressed the most frequently asked questions across the state. The addenda materials were created and/or reviewed by the Regional Coordinators at an implementation issues meeting held in late fall. Concurrent with the March, 1993, dissemination of this addenda, regional Cluster Leader Meetings were held to review the new materials and clarify other portfolio issues. A KET telecast was also conducted to assist Cluster Leaders in sharing the new materials with their cluster members.

ASSESSMENT OF NONCOGNITIVE INDICATORS

In addition to the cognitive indicators used to measure school success, KERA mandates that noncognitive indicators be factored into the accountability index. A subcontractor for the assessment, Far West Laboratories, collects and analyzes the noncognitive data. The noncognitive indicators represent one sixth of the overall school accountability index in the first accountability cycle, and are composed of several indicators, not all of which are used at each grade level.

- **Attendance Rate**, defined as the ratio of aggregate attendance to the sum of aggregate attendance and aggregate absence (used at all grade levels);
- **Retention Rate**, defined as the ratio of number retained to total membership (used at all grades' 4-12 levels);
- **Dropout Rate**, defined as the ratio of number of dropouts to total membership (used at grades' 7 and 12 only); and
- **Successful Transition Rate**, defined as the ratio of number of "successful" high school graduates to the total number of graduates. As detailed in 1991 by the Department of Education and published in *Guidelines for Transition to Adult Life*, "successful transition to adult life" is identified by such criteria as full-time enrollment in post-secondary school, employment in a non-temporary position at least 30 hours per week, active military status, or a combination of these situations totaling at least 30 hours per week. Transition is assessed by school administrators who must verify "successful" transition with either the student or a member of the student's family (used in grade 12 only).

ACCURACY OF NONCOGNITIVE DATA. Noncognitive data from 1991-92 and 1992-93 were examined in a study of the accuracy of the noncognitive indicators and the procedures used to collect the data. (A full discussion of reliability issues for KIRIS noncognitive indicators can be found in Appendix F of the 1992-93 Technical Report.) In general, the analyses focused on:

- the thoroughness of the data-collection procedures including check-in and auditing strategies;
- stability within years, as indicated by modifications of submitted data following review by school and district staff of their original submissions;
- stability across years, as indicated by comparisons of 1992-93 data to previous year's data; and
- interpretability of the data across indicators at the various grade and statewide levels

While these analyses found the procedures used to collect and report the KERA noncognitive indicators to be thorough and reliable, they suggested that KDE expand the number of options for districts to submit data, emphasizing methods that take advantage of improving local technology capabilities, and at the same time standardize noncognitive data-collection procedures and timelines and eliminate duplication of effort. In addition, findings suggested that KDE should Increase the training of local staff on the modified definition of dropouts through memoranda, meetings, and other techniques, as well as consider a pilot study examining the accuracy of local data-collection procedures across the indicators.

CHAPTER 8
SCORING

OPEN-RESPONSE QUESTIONS AND ON-DEMAND WRITING

As in 1991-92 and 1992-93, the 1993-94 KIRIS open-response questions, writing samples, and performance events at grades 4, 8 and 12 required hand-scoring by Advanced Systems personnel. Portfolios in writing and mathematics as well as alternate portfolio were hand-scored by Kentucky teachers. While the processes of selecting and training scorers, reading and scoring papers, and monitoring scoring remained very similar to those carried out in previous years, as described in the 1991-92 Technical Report, these procedures are described below in detail. Modifications in training materials for scorers and teachers are described as well.

STAFFING. With a greater number of items to score within the testing and reporting schedule, more scorers were needed in 1993-94 and 1992-93 than in 1991-92. Levels of staffing are listed in Table 8-1. The table also shows the percentages of scorers at each grade level who participated in a previous year's scoring (repeat scorers), as well as the number of training leaders.

TABLE 8-1
NUMBER OF SCORERS AND TRAINING LEADERS AT EACH GRADE

Grade	1991-92		1992-93			1993-94*		
	Scorers	Training Leaders	%Repeat Scorers	Scorers	Training Leaders	%Repeat Scorers	Scorers	Training Leaders
4	98	11	93	129	14	73	130	11
8	98	10	68	156	14	73	130	11
12	82	10	95	118	14	73	130	11

* The same scorers were used for grades 4, 8 and 12 in 1993-94

QUALIFICATIONS. Table 8-2 shows qualifications and basic demographics for scorers in each testing year.

TABLE 8-2
PROFILE OF SCORER QUALIFICATIONS AND DEMOGRAPHICS

Background		Number of Scorers						
		1991-92			1992-93			1993-94
		Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12	All Grades
Education	Degrees beyond the Baccalaureate	24	24	24	18	32	13	22
	Bachelor's Degree	64	57	48	93	104	93	80
	Associate's Degree	5	5	5	10	3	7	8
	Two-year college study or equivalent and previous scoring experience	5	12	5	8	17	5	20*
Experience	Administration	4	4	4	1	2	0	2
	Teaching	27	20	26	38	126	25	40
	Aide or substitute	11	11	11	4	5	4	20
	Other/non-school (journalist, accountant, etc.)	56	63	41	86	23	89	68
Demographics	Male	32	32	20	37	53	35	53
	Female	66	66	62	92	103	83	77
	Black	2	2	2	3	3	3	1
	White	94	93	80	122	153	115	129
	Other	2	3	0	4	0	0	0

* Two scorers were honor students just completing their first year of college. They received extra monitoring in the scoring process.

SCORER TRAINING. The training of scoring staff was completed in two phases: training of the chief readers and training leaders, followed by training of the scorers.

The chief reader and the training leaders for open-response questions in each content area met with the Advanced Systems' test developer responsible for that domain. The developer, as a member of the Kentucky advisory committee, presented the Kentucky objectives, standards, and background information necessary to understand the objectives being measured. Each group also reviewed the framework of the scoring guide and the language pertinent to the standard.

After this introduction, the combined group read hundreds of student responses and selected benchmark exemplars — papers which typify each score point in the scoring guide. Once the benchmarks were established, the training leaders continued the preparation. They identified a second set of benchmark papers, similar to the first training pack, and additional examples of student responses that represented a range within each score point. Throughout this process, development staff was available to discuss concerns presented by the training leaders and answer any questions that they might have. Before training packs were reproduced, the training leader met with the developer for a final review of the training materials.

Advanced Systems' development staff observed the initial sessions when the training leaders presented the standards to the scorers and were available to provide additional clarification if needed. The training leader then completed the training independently. The training leader and developer consulted as needed throughout scoring; at the end of the project, development staff and leaders met to share information about the process and to offer suggestions and comments for future improvement. Throughout the scoring period, training leaders read student responses as they checked the reliability of each scorer. As they read, training leaders also selected recaliberation papers that were reviewed with scorers at the beginning of each day to assure that drift from scoring standards did not occur.

Scorers for each content area were selected for their content expertise and were trained by the training leaders. The training leaders first presented background information and an explanation of the scoring guide. The first set of training papers — the benchmark training pack — was used to clarify the language of the scoring guide; each score point was illustrated by an exemplar. This pack became the reference set used throughout scoring. Scorers were instructed to review the language of the guide regularly as they read actual student responses.

The first training pack was similar to the benchmark pack, but papers were in random rather than sequential order by score point. A second training pack was designed to teach scorers how to identify a range within each score point. After discussing the papers in each pack, scorers were asked to assign scores independently to another set of papers. The scores were compared to those assigned by the training leaders and developers. The final training pack consisted of unique student responses to the question or writing prompt.

As a final qualifying step, scorers were instructed to score six papers. The training leaders checked each person's responses for accuracy; those who achieved success on four of the papers in assigning the appropriate score point began to score actual papers. Those who needed further training worked with the training leader until they were able to achieve the two -thirds rate of agreement required to qualify. All scorers hired for the project qualified in this process.

Training papers in 1991-92 included a benchmark set and two types of training packs. For 1992-93 and 1993-94 scorings, the training papers were modified. The benchmark pack, as in 1991-92, expanded on the language of the scoring guide and provided exemplar responses to illustrate each score point. One training pack was used in 1992-93 and 1993-94, as opposed to two in 1991-92. The training pack for 1992-93 and 1993-94 included several student samples in random score order, some clearly similar to benchmark samples, some different. Scorers' work on this set was compared to those scores assigned by the training leaders and developers to determine whether they were ready to begin actual scoring.

Training for the scoring of writing was similar in all ways to that of the open-response questions except for the initial step. The chief reader and training leaders for writing met with the Advanced Systems writing consultant and the Kentucky writing committee to establish the standards and select the benchmark papers. Other procedures were the same for the rest of the preparation and training process.

RELIABILITY OF SCORING. As in 1991-92 and 1992-93, scoring of the 1993-94 open-response tests was monitored in two ways. The first was ongoing; training leaders constantly moved from scorer to scorer, re-reading samples of each scorer's work. Each training leader read approximately two papers per scorer each morning and afternoon. The second monitoring procedure was a rescoring of all readers, including the training leaders. Two percent of the total responses was rescored in this way. No adjustments were made, however, to discrepant scores. Less than one percent of the scores subject to rescoring were found to be discrepant by more than one point, according to the verification data collected. The reliability of scoring — percentages of exact matches between scores and percentages of scores assigned within adjacent points — is shown in Table 8-3.

TABLE 8-3
INTER-RATER RELIABILITY IN SCORING OF OPEN-RESPONSE QUESTIONS

Grade	Percentage of Exact Agreement				Percentage within 1 Score Point			
	4	8	12	Total	4	8	12	Total
Reading	81.3	85.2	85.1	84.8	99.2	99.8	99.6	99.6
Math	86.5	93.4	87.2	90.3	98.9	99.8	99.5	99.6
Science	82.8	84.9	81.1	82.9	99.1	99.5	99.2	99.3
Social Studies	84.4	82.4	83.0	82.9	99.5	99.9	99.7	99.8
Total	83.7	86.4	84.0	85.1	99.2	99.8	99.5	99.6

As another check on scoring reliability, results of first and second scorings were compared. Table 8-4 shows correlations between the two scorings across the common open-response questions.

TABLE 8-4
STUDENT-LEVEL INTER-RATER CORRELATIONS

Subject	Grade 4	Grade 8	Grade 12
Reading	.90	.92	.91
Mathematics	.95	.98	.93
Science	.92	.92	.90
Social Studies	.91	.90	.89

PERFORMANCE EVENTS

When the team of scorers completed the scoring of grade 8 open-response questions, the final set of open-response materials to be scored, many were retained to score the KIRIS performance events for all grade levels. The training procedures and score monitoring by the training leader were the same as for the open-response tests. That is, Advanced Systems' test development staff and content advisory committees were responsible for writing scoring guides for each task and then Advanced Systems staff worked with the training leaders to select student responses that illustrated score points in the guides. The student responses were assembled into packets that were used to train the group of scorers. Each training leader reviewed at least two responses per scorer each morning and afternoon as a way of monitoring scorer reliability. Scores for performance events at grade 4 were assigned holistically (a single score per event). At grades 8 and 12, each performance event score was the sum of multiple scored items within the task.

Because performance materials were packaged by school and content area, managing the paper flow was different than for the open-response questions or writing. Each school's materials were counted and the count was recorded on a header sheet identifying the number of students completing the tasks. Before the student papers were distributed at a table in a scoring cluster, the count was verified. After scoring was completed, materials were counted again before being repackaged in their original containers.

MATHEMATICS AND ALTERNATE PORTFOLIOS

Mathematics portfolios and alternate portfolios were added to the KIRIS assessment in 1992-93. These portfolios will be examined in the Technical Manual for the second accountability cycle.

WRITING PORTFOLIOS

In 1991-92, a five-point holistic scoring method was selected, after which a scoring guide was developed. Teachers evaluated student work on a five point scale, with additional analytic notations indicating specific needs and commendations. While the initially-selected criteria for evaluation and content remained the same for 1992-93, the five-point holistic scoring method was collapsed into a four-point holistic method. Specifically, the two lowest points on the scale were combined, both indicating a novice performance level. No changes were made in 1993-94.

In 1992-93, after all writing portfolios were scored by teachers, DACs identified a sample of five of each teacher's portfolios to be rescored within the district. Teachers then make copies of the selected portfolios, which the DACs redistribute to other teachers in the district for blind rescoring. After the rescore data was provided to Advanced Systems, discrepant teachers were identified on basis of mean and absolute mean differences between their original scores and the rescores. All portfolios from the discrepant teacher were then copied and rescored during the summer at sessions conducted by the Department of Education.

In 1993-94, within-district rescoring gave way to a broad-based approach to giving feedback at the school level. In this process, a sample of portfolios from 50% of all Kentucky schools was rescored at a summer scoring session. Samples of portfolios from the remaining schools will be rescored next summer, thereby providing appropriate feedback to all schools in the scoring of portfolios. Chapter 12 describes several studies of writing portfolio scoring.

CHAPTER 9
EQUATING

INTRODUCTION

Because schools are to be evaluated on the basis of their improvement over time, test results must be linked from year to year. By placing the results of the 1993-94 open-response questions and performance assessments on the same scale used in 1991-92 and 1992-93, it is reasonable to use 1992-93 and 1993-94 scores to judge progress against the standard of the 1991-92 baseline.

EQUATING OPEN-RESPONSE TESTS

PRELIMINARY CONSIDERATIONS. Given the number of open-response items present in the 1991-92 test and the release of common items every year, it was not feasible to make a direct comparison of the 1991-92 and 1993-94 open-response tests; the two tests shared too few items to make an effective equating possible. Because performance-level assignments in 1992-93 had been placed on the 1991-92 scale in a previous equating, it was judged sufficient to equate the 1993-94 open-response test results to the 1992-93 results. However, changes in the scoring of open-response items in 1993-94 and in the placement and number of items used in 1992-93 necessitated the consideration of several modifications to the equating approach used previously, described in the 1992-93 Technical Report.

Reused 1992-93 items on the 1993-94 test form the basis of this equating. Because common items from 1992-93 were released and therefore unavailable for 1993-94, all of the reused items were matrix questions on the 1992-93 assessment, positioned among the final two questions on each form. Some of these reused items served as common items in 1993-94, where they were among the first five questions on every form. Because relative placement could significantly affect performance on specific questions, year-to-year differences in performance on those questions might reflect the influence of extraneous factors rather than changes in student ability. Of particular concern is the possibility that students taking items at the end of a test might be fatigued relative to students taking the same items at the beginning of a test.

To avoid the possibility of item placement having an deleterious impact on the equating, it was decided that matrix items would serve as the sole basis for the equating. Common items could be dropped from consideration in this case because of the large number of re-used matrix items.[1] This approach was preferred to the one used in equating 1992-93 results to 1991-92, in which common items were used to link tests. In this approach individual items were deleted from the equating solely on the basis of statistical evidence of differences in difficulty across the two testing years.

The more difficult issue to consider is the revised scoring procedure used in 1993-94. As described in chapter 4, the 1993-94 open-response items were scored using a different scoring system than the 1992-93 items. Specifically, the 1992-93 category designated '1' (indicating a minimal or off-topic response) was divided into two 1993-94 categories: '1', indicating a minimal response; and '0', indicating an off-topic response. The remaining categories, 2, 3, and 4, (as well as the designation of 'Blank' for no response) retained their assigned meanings. Because of changes in the definition of the 0 and

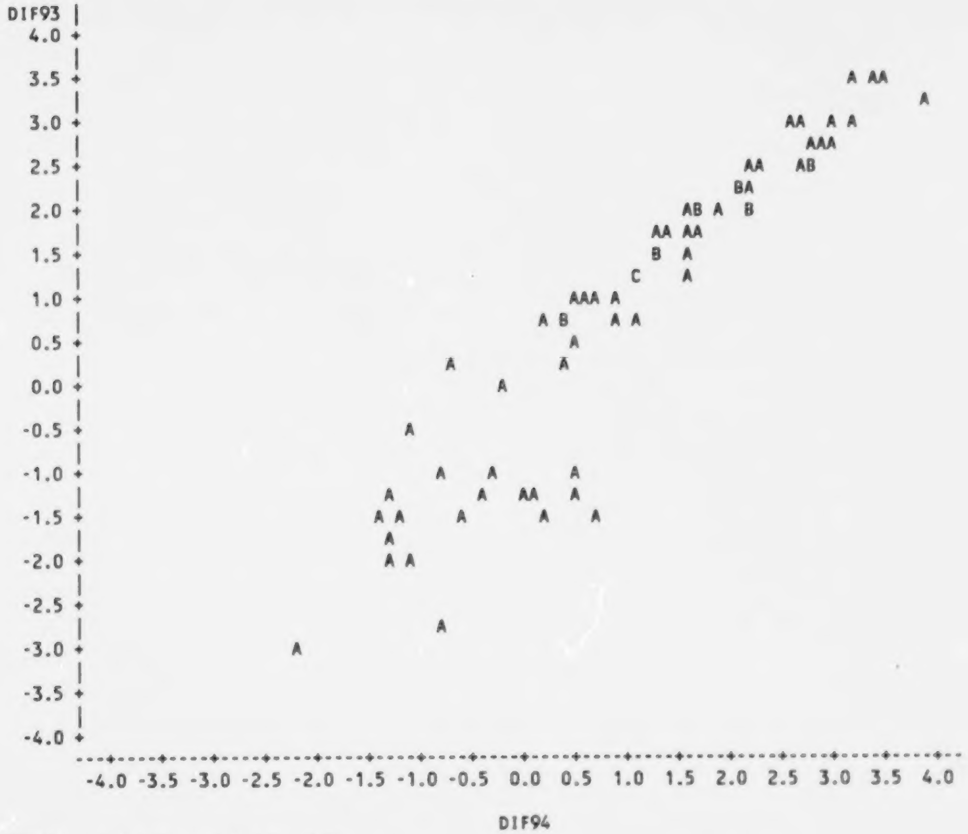
---

[1] This approach was used for every grade and content area combination except grade 4 Science, in which only three matrix items were re-used as matrix items. In this case, the equating was based on the three matrix items and the five common items.

1 score points, there was no question that scoring at these levels would not be consistent from 1992-93 to 1993-94. What remained unknown, however, was whether (and the extent to which) this change would affect scoring at other points along the scale.

To examine this question, item difficulties for 1993-94 and 1992-93 administrations of the same items were plotted including and excluding 0/1 item difficulties. Figure 9-1 and Figure 9-2 present these respective plots for grade 12 Mathematics items. Whereas the 0/1 difficulties varied greatly, as seen in the lower left-hand corner of Figure 9-1, the remaining difficulties demonstrated a nearly linear relationship. Thus, the results of item calibration suggested that the meaning of the 2, 3 and 4 score points had been maintained successfully. For this reason, only the 1/2, 2/3 and 3/4 item difficulties were used in the equating.

FIGURE 9-1
DIFFICULTIES FOR 1993-94 MATHEMATICS ITEMS RE-USED FROM 1993-92, GRADE 12
PLOT OF 1992-93 DIFFICULTIES (DIF93) WITH 1993-94 DIFFICULTIES (DIF94)

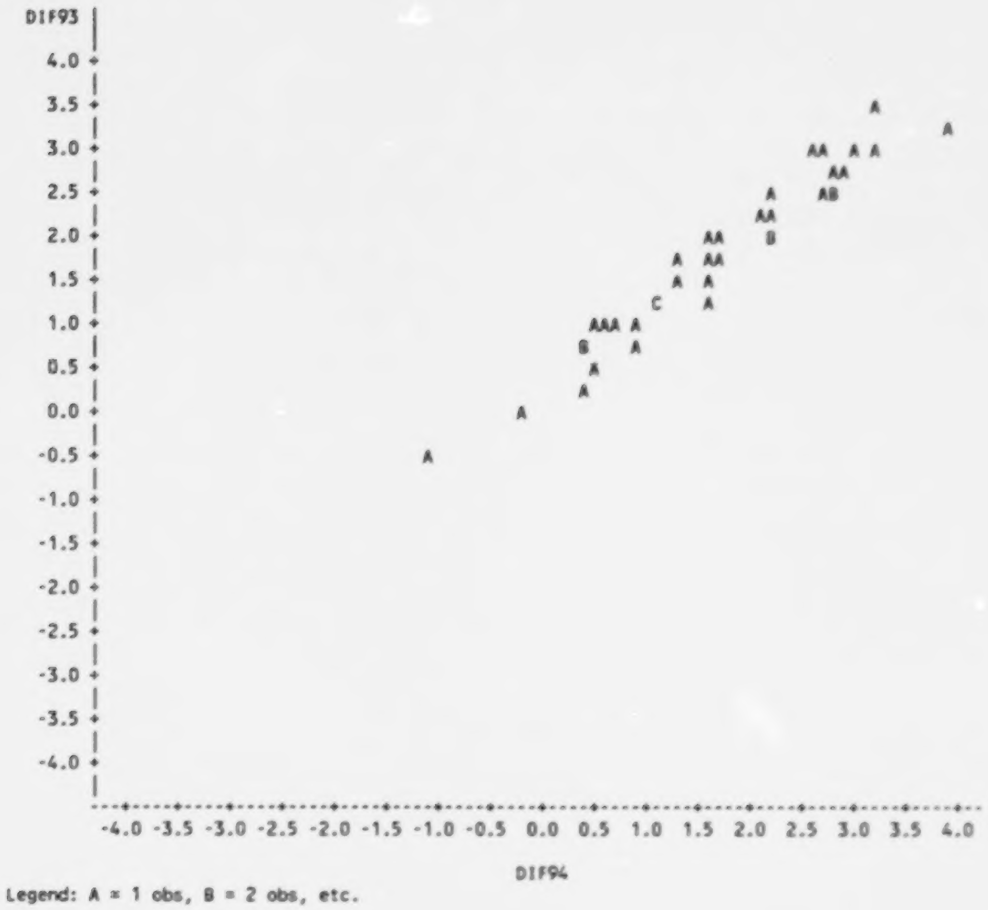


Legend: A = 1 obs, B = 2 obs, etc.

FIGURE 9-2

DIFFICULTIES FOR MATRIX-SAMPLED 1993-94 MATHEMATICS ITEMS RE-USED FROM 1993-92, GRADE 12

PLOT OF 1992-93 DIFFICULTIES (DIF93) WITH 1993-94 DIFFICULTIES (DIF94)



Note that item difficulties greater than 4.0 do not appear in Figures 9-1 and 9-2, and that Figure 9-2 includes item difficulties only for the 14 open-response items presented in matrix sampling in 1992-93 and 1993-94. For these reasons, only 68 points are plotted in Figure 9-1, and only 38 points appear in 9-2.

PROCEDURES. The following paragraphs describe the methods used to equate the 1993-94 open-response test results to the 1992-93 open-response test results. Because the 1992-93 open-response test results were previously equated to the 1991-92 results, the process described below effectively placed the 1993-94 results on the original 1991-92 scale. MULTILOG® (1991), software specifically designed for items with multiple responses, was used to facilitate item analysis and scoring of the open-response items within the framework of Item Response Theory (IRT).

The steps in the equating process are listed below. A detailed description of each step follows.

1. Scale 1993-94 open-response items.
2. Plot item difficulty parameters for linking items; eliminate questionable items from equating.
3. Put 1993-94 item parameters on 1992-93 scale.
4. Score students on the 1993-94 open-response items.
5. Compare matrix and common item-based performance level estimates to determine adjustment to cut scores.
6. Examine rescored data to determine adjustment to ability estimates.
7. Calculate **final** 1993-94 abilities, applying ability estimate adjustments.
8. Assign 1993-94 students to performance levels using adjusted cut scores.

1. Scale 1993-94 open-response items.

The first step in the equating process was to scale the 1993-94 open-response items. Responses from all tested students having no blank responses in a given content area were used in the item calibrations. Samejima's two-parameter graded model for ordered responses (1969) was used to estimate the item parameters (difficulties and discriminations) for the five common open-response items for each of the four content areas. Then, fixing the item parameters for the common items, each of the twelve forms of the assessment at each grade level was calibrated separately estimating the item parameters for the pairs of matrix-sampled items[2].

2. Plot item difficulty parameters for linking items.

Many of the 1993-94 items had been administered in the 1992-93 assessments, and could therefore be used to link scores between the two years. Table 9-1 shows the number of 1993-94 common and matrix items by grade and content area which had been given in the 1992-93 assessment. Matrix items are noted first, followed by common items in parentheses. (For example, 14 out of the 24 matrix-sampled items from the 1992-93 reading assessment were included in the grade 4, 1993-94 reading assessment; five appeared as common items and nine as matrix-sampled.) Overall, four or five of the common open-response items for each content area on the 1993-94 assessment were unreleased matrix-sampled items from the 1992-93 administration. The number of unreleased 1992-93 matrix-sampled items in each content area administered again as matrix items in the 1993-94 assessment ranged from 3 items in grade 4 Science to 19 items in grade 12 Science. (Specific item locations in each year's assessment are presented in Appendix 9-1.)

---

[2] Although form three of grade 8 Social Studies contained one item rather than two, matrix items will be referred to as occurring in pairs on each form for the sake of brevity.

TABLE 9-1
NUMBER OF 1993-94 MATRIX AND COMMON ITEMS ADMINISTERED AS MATRIX ITEMS IN 1992-93

Grade Level	Grade 4	Grade 8	Grade 12
Reading	9 (5)	13 (5)	11 (5)
Mathematics	10 (5)	17 (5)	14 (4)
Science	3 (5)	16 (5)	19 (4)
Social Studies	11 (4)	17 (5)	17 (5)

The 1993-94 item difficulty parameters for the linking items for each content area were plotted against their respective 1992-93 item difficulties, as determined in the previous year's equating, detailed in the 1992-93 Technical Manual. Based on the four score points used in this analysis (combined zero and 1 scores, 2, 3, and 4), each item provided three difficulty estimates, one at the 0,1/2 score point, one at the 2/3 score point, and one at the 3/4 score point. For example, the 9 linking reading items for grade 4 resulted in 27 item difficulty parameters. These location parameters, or thresholds, specify the point on the ability (Θ) scale at which a person has a 50% chance of responding in some higher category than the one to which the threshold belongs.[3] Through a visual examination of the plotted difficulties, any item difficulty that deviated from the 45° line drawn through all points was identified as an outlier and excluded from the equating analysis. Table 9-2 presents the number of item difficulties identified as outliers in each content area at each grade level. Figure 9-3 shows the plot of difficulties for the grade 4 reading items. The two item difficulties in parentheses represent the outliers identified. After all outliers were identified, the mean and standard deviation of the remaining item difficulties at the selected score points were calculated for each year of testing.

TABLE 9-2
NUMBER OF ITEM DIFFICULTIES IDENTIFIED AS OUTLIERS

SUBJECT	Grade		
	4	8	12
Reading	2	0	0
Mathematics	2	4	2
Science	2	5	3
Social Studies	2	2	0

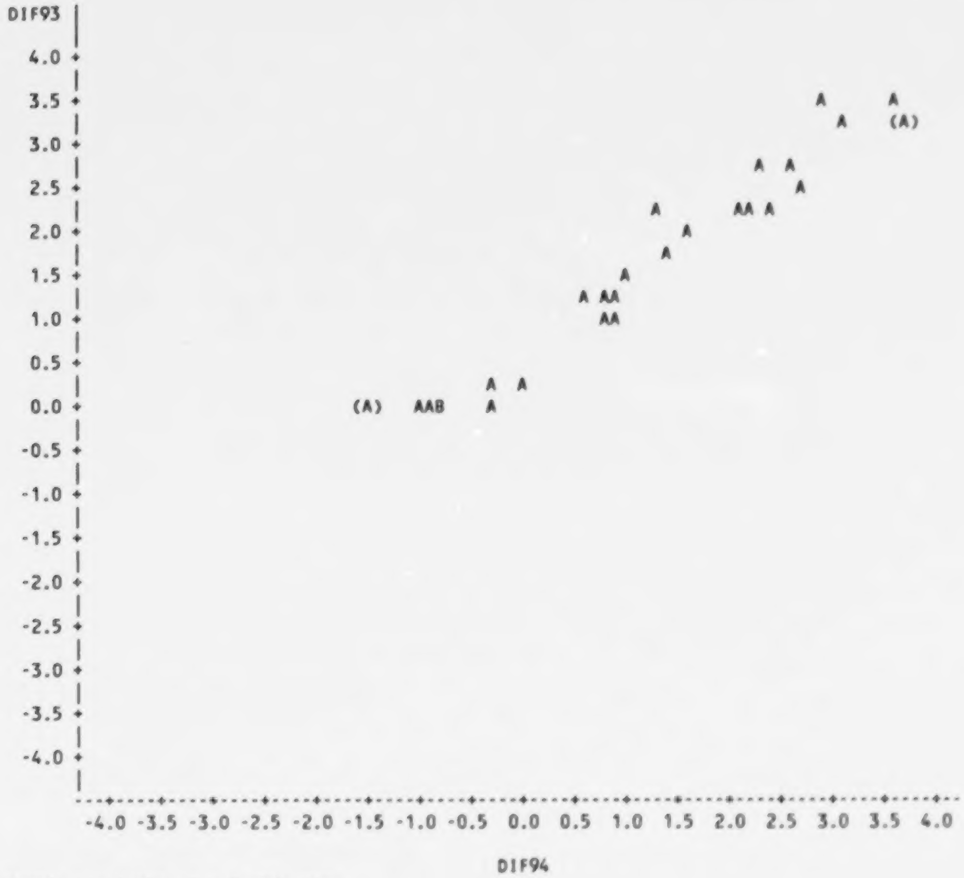
---

[3] Thissen, David, *MULTILOG User's Guide*, Scientific Software, Inc., 1991

FIGURE 9-3

DIFFICULTIES FOR MATRIX-SAMPLED 1993-94 READING ITEMS RE-USED FROM 1993-92, GRADE 4

PLOT OF 1992-93 DIFFICULTIES (DIF93) WITH 1993-94 DIFFICULTIES (DIF94), WITH OUTLIERS NOTED



Legend: A = 1 obs, B = 2 obs, etc.

3. Put 1993-94 item parameters on 1992-93 scale.

The 1993-94 item parameters for all items, even those not included in the equating analysis, had to be adjusted by putting them on the 1992-93 scale. This includes the item parameter estimates for the 0/1 and 1/2 score points of the linking items as well as all parameter estimates associated with the common and non-linking matrix items. Steps outlined by Hambleton and Swaminathan (1985) were followed to equate parameters generated from a two-parameter model. Using the mean (mn) and standard deviation (sd) of the item difficulties from the 1992-93 ("92") assessment and the 1993-94 ("93") assessment, the first step was to calculate coefficients α and β.

$$\alpha=(sd_{92}/sd_{93}) \qquad \beta=mn_{92}-\alpha(mn_{93})$$

Table 9-3 presents the mean and standard deviation of the item difficulties, as well as α and β for each grade and content area.

TABLE 9-3
ITEM DIFFICULTY STATISTICS AND ADJUSTMENT COEFFICIENTS

Grade	Subject	mn_{92}	mn_{93}	sd_{92}	sd_{93}	α	β
4	Reading	1.5532	1.1804	1.1749	1.3725	0.856	0.5428
	Mathematics	1.3282	1.0743	0.9599	1.0839	0.8856	0.3768
	Science	0.9045	0.843	1.2449	1.3859	0.8983	0.1472
	Social Studies	1.161	1.0335	1.3369	1.3601	0.983	0.145
8	Reading	1.2336	0.8431	1.1811	1.2303	0.96	0.4242
	Mathematics	1.536	1.4291	0.9407	1.1016	0.8539	0.3158
	Science	1.6285	1.6717	1.3449	1.4882	0.9037	0.1178
	Social Studies	1.8064	1.5355	1.2703	1.3423	0.9463	0.3533
12	Reading	0.9988	0.7997	1.0644	1.2769	0.8336	0.3322
	Mathematics	2.0103	1.9138	1.2093	1.4046	0.861	0.3626
	Science	2.0558	2.0162	1.3527	1.5725	0.8602	0.3214
	Social Studies	1.6186	1.3849	1.2786	1.4991	0.8529	0.4375

The 1993-94 discrimination item parameters (a) were then adjusted by the following equation.

$$a'_{93}=(a_{93}/\alpha)$$

The 1993-94 difficulty item parameters (b) were adjusted by the following equation.

$$b'_{93}=\alpha(b_{93})+\beta$$

As the last step in the process, the **final** 1993-94 item difficulty parameters (b''_{93}), put on the 1991-1992 scale for items on both the 1991-1992 and 1992-1993 assessments, were calculated by averaging the 1992-93 item difficulty parameters (b_{92}) with the adjusted 1993-94 item difficulty parameters (b'_{93}).

$$b''_{93}=\frac{b_{92}+b'_{93}}{2}$$

Similarly, the **final** 1993-94 item discrimination parameters (a''_{93}) for those items were calculated by averaging the 1992-93 item discrimination parameters (a_{92}) with the adjusted 1993-94 item discrimination parameters (a'_{93}).

$$a''_{93}=\frac{a_{92}+a'_{93}}{2}$$

4. Score students on the 1993-94 open-response items.

Using the final adjusted 1993-94 item parameters for the common open-response items, estimates of student abilities were calculated using Samejima's graded model. Ability estimates were also calculated, by form, on the pairs of matrix-sampled open-response items.

5. Compare matrix- and common item-based performance level estimates.

Using the final adjusted 1992-93 ability estimate to performance level translation table, the ability estimates computed for students on the 1993-94 common and matrix items were translated into performance levels. Whereas scores from previous years had yielded equal percentages across matrix and common items, there was great variability in agreement by grade and content area. In the most discrepant case, grade 12 mathematics, matrix-based ability estimates yielded a distribution of performance levels in which about half the students were classified as *Novice*, yet only about a third of the students were classified as *Novice* on the basis of the common items. To remedy this discrepancy, matrix item-based performance level cut scores on the ability estimate scale were adjusted to produce percentages of *Novice*, *Apprentice*, *Proficient* and *Distinguished* students equivalent to the common item-based percentages. Because the five common items produced ability estimates containing less error than the two item, matrix-based estimates, it was reasoned that the common item-based performance level percentages more accurately reflected the underlying ability of Kentucky students. Therefore, the matrix-based cut score points were adjusted rather than the common item-based cut score points. Table 9-4 examines the impact of this adjustment. If this adjustment was not appropriate, its effect would have been to artificially raise the statewide accountability index at each grade by about one and one-half points across the four content areas.

TABLE 9-4
IMPACT OF ADJUSTING CUT SCORES USED TO ASSIGN PERFORMANCE LEVELS ON MATRIX-SAMPLED ITEMS

Grade	Content Area	Statewide Adjustment to Performance Levels Based on Matrix-Sampled Items, in Percent				Statewide Adjustment to Index, on 0 to 140	
		Novice	Apprentice	Proficient	Distinguished	Content Area Index Based on Matrix Items	Total Index
4	Reading	-5.5	-0.3	4.9	0.9	6.0	0.5
	Mathematics	-7.0	3.7	1.0	2.3	5.7	0.4
	Science	-1.8	0.3	0.4	1.0	2.0	0.1
	Social Studies	-4.6	-0.1	3.6	1.0	5.0	0.3
8	Reading	-2.4	-0.5	2.5	0.5	2.9	0.2
	Mathematics	-10.3	6.3	2.8	1.1	6.9	0.5
	Science	-4.8	3.8	0.8	0.3	2.6	0.2
	Social Studies	-9.6	6.7	1.1	1.9	6.3	0.4
12	Reading	-5.8	4.8	0.5	0.5	3.1	0.3
	Mathematics	-15.8	11.5	1.9	2.4	9.9	0.7
	Science	-6.4	5.5	0.3	0.5	3.2	0.2
	Social Studies	-4.6	-1.7	5.2	1.2	6.1	0.4

6. Examine rescored data to determine adjustment to ability estimates.

It was important to ensure that any change in ability across the two assessment cycles was a reflection of a change in the students' abilities and not due to any difference in the scoring of the open-response items. Though the training of the scorers was reasonably consistent across years, different scorers can score differently, possibly scoring higher or lower than their counterparts. This concern is especially great when there has been a change in the scoring rubric, a situation that limits consideration of year-to-year score comparability to the unchanged scoring categories.

To determine if there was any change in scoring across years, a sample of student responses to the linking items from the 1992-93 assessment, 50 responses at each score point, was rescored by 1993-94 scorers. Scores from the 1993-94 scores were compared with original scores from 1992-93. The rescoring data was used to recast the original 1992-93 results. That is, the conditional probabilities generated from the rescoring results were used to compute posterior distributions to obtain a total for each score point. As Table 9-5 shows (using grade 12 data), the percentages of students receiving a score of 2, 3, or 4 is consistent whether one looks at the original 1992-93 scores or the recast 1992-93 scores. The scoring did change where it was expected to change: the scoring of papers receiving a score of 0 or 1.

TABLE 9-5
CROSSTABULATION OF 1993-94 RESCORING OF GRADE 12 READING RESPONSES WITH SCORES ORIGINALLY ASSIGNED IN 1992-93, IN PERCENT

		Score assigned in 1993-94					Percent assigned to score category in 1992-93
		0/B	1	2	3	4	
Score Assigned in 1992-93	0	8	0	0	0	0	8
	1	6	24	11	1	0	42
	2	0	12	18	5	1	36
	3	0	1	5	5	1	12
	4	0	0	1	1	1	3
Percent assigned to score category in 1993-94		14	37	35	12	3	

On the basis of these results, and similar results across content areas and grades, it was determined that any possible scoring differences from 1992-93 to 1993-94 were not major enough to warrant any changes to the 1993-94 data. Therefore, the only change made to 1993-94 abilities was to bring the 1993-94 scores into line with the 1991-92 scoring standards. This was accomplished by making the same 1991-92 rescoring adjustment to the 1993-94 scores that was made to the 1992-93 scores. These adjustments were taken from the previous year's calculations, and added to the 1993-94 abilities as the rescoring adjustment. Table 9-6 presents the rescoring adjustments.

TABLE 9-6
RESCORING ADJUSTMENTS FOR 1993-94 ABILITIES

Subject	Grade 4	Grade 8	Grade 12
Reading	0.0724	0.0666	0.0767
Mathematics	0.0482	0.0152	-0.0078
Science	0.082	0.0276	-0.0054
Social Studies	0.0597	-0.0006	-0.0324

7. Calculate final 1993-94 abilities.

For each content area, the 1993-94 ability estimates calculated in step 5 were corrected by the rescoring adjustments. Non-tested students were identified and assigned the lowest ability estimate, accounting for all included students.

8. Assign 1993-94 students to performance levels using cut scores.

It was discovered that for some forms of the test, students receiving a perfect score on the two matrix-sampled items (a total of 8 points) were identified as only performing at the *Proficient*, not *Distinguished*, performance level. In order to allow every student an opportunity to obtain a Distinguished rating, a perfect score on any test was deemed *Distinguished*. This policy-based decision meant that the ability cut-offs for each performance level from the scoring of the 1992-93 common open-response items were applied to the 1993-94 common item-based ability estimates for each content area. The adjusted cut-scores calculated in step 5 were applied to the 1993-94 final ability estimates based on matrix-sampled items for each content area. After performance levels were assigned, the frequencies of the performance levels by total raw score points on the five common open-response items were examined. Where possible, each common total score point was assigned to a single performance level. In cases where reversals occurred, e.g., instances were a raw score of 7 would be classified as *Apprentice* and a raw score of 8 would be classified as *Novice*, classification based on raw score was followed. Such reversals occurred on less than 2% of all occasions, on average. Appendix 9-2 gives cut scores used to assign performance levels based on common item total raw scores.

When using Item Response Theory (IRT), it is difficult to accurately estimate a student's ability (θ) when the student receives a perfect score on the test. When assigning proficiency levels based on the matrix-sampled items, it was discovered that for some forms of the test, students receiving a perfect score on the two matrix-sampled items (a total of 8 points) were identified as only performing at the *Proficient*, not *Distinguished* rating, a perfect score on any test was deemed *Distinguished*. This policy-based decision is consistent with practice in other testing programs such as the SAT.

Therefore, perfect scores on the matrix-sampled items always resulted in a classification of *Distinguished*. Table 9-7 shows the percentage of students classified as *Distinguished* based on the original IRT estimates for the matrix-sampled items, and then based on the decision to change all perfect scores on the matrix-sampled items to a classification of *Distinguished*. The percentage of students classified as *Distinguished* based on the common items is also provided for comparison purposes. No differences are apparent between the original and final assignments presented in Table 9-7.

TABLE 9-7
PERCENTAGES OF 1993-94 STUDENTS CLASSIFIED AS DISTINGUISHED ON OPEN-RESPONSE ITEMS

Grade	Subject	Based On Common Items	Based On Matrix-Sampled Items	
			Original Assignment	Final Assignment
4	Reading	1	1	1
	Mathematics	5	5	5
	Science	1	1	1
	Social Studies	1	1	1
8	Reading	1	1	1
	Mathematics	8	8	8
	Science	0	0	0
	Social Studies	2	2	2
12	Reading	2	2	2
	Mathematics	6	6	6
	Science	1	1	1
	Social Studies	2	2	2

PERFORMANCE EVENTS EQUATING

PRELIMINARY CONSIDERATIONS. Because all performance events are to released each year no tasks from previous years were included in the 1993-94 assessment. Therefore, it was impossible equate the performance events results using the same common item equating model used in the equating of the open response tests. The most feasible option available was to equate the performance events using the **common person equating model.**

Although the decision to use common person equating was uncomplicated, the selection of an actual equating design presented more challenges. In 1992-93 several alternative methods were considered which can be classified into three basic approaches:

- Administer events from the previous year to a sample of grade 4, 8, and 12 students Kentucky students in conjunction with the performance events testing.
- Administer events from the current year and previous year to off-grade Kentucky students (e.g. grade 3, 5, 7, or 11) in conjunction with the performance events testing at grades 4, 8, and 12.
- Administer events from the current year and previous year to grade 4, 8, and 12 students outside of Kentucky.

Each of these approaches has its own advantages and disadvantages - from both administrative and psychometric viewpoints. After consultation with the Technical Advisory Committee the approaches involving the testing of Kentucky students were rejected primarily because of the wide distribution of previous performance events throughout the state. It seemed likely that it would be difficult to find a representative sample of students who had not been exposed to and/or spent a considerable amount of time reviewing the released events.

The particulars of the design involved administering performance events from the current year and the previous year to the same sample of students. Each student would complete two performance tasks - one from each year of testing. This procedure would be followed for each grade (4, 8, and 12) and content area (mathematics, science, social studies) combination. The theory behind the design is as follows:

- Based on their performance on the performance events from the previous year it would be possible to determine the distribution of students across performance levels.
- Having determined the distribution of students across performance levels it would be possible to establish cut scores on the performance events from the current year that would result in the same distribution of students.

A key component of any common person equating approach is that groups of students completing each set of tasks are equivalent in ability. Ideally, this is accomplished by using a single group. That is, the same people complete both sets of tasks. Unfortunately, in conducting the performance events equating it was not possible to require that the students complete two tasks from the same content area. Schools that volunteered to participate in the study wished to expose each student to tasks from as many content areas as possible. In the absence of any additional data on the participating students, it was also not possible to guarantee that the groups completing each set of tasks were equivalent. Left with no alternative, however, the equating was conducted based on the assumption that the groups were equivalent.

MODIFICATIONS FOR 1993-94. The purpose of the equating is to make it possible to judge progress against the standard of the initial baseline. As implied in the prior section on the equating of open-response tests it is desirable to make a direct comparison of the 1991-92 and 1993-94 tests. Although that was not feasible for the open-response tests, there were no obstacles to accomplishing this comparison for the performance events tests. Therefore, the sets of events administered during the equating study were from the initial year of testing and the current year of testing. For mathematics, science and social studies those were the 1991-92 and 1993-94 tests. In those content areas no performance events from the 1992-93 test were administered in the equating study.

An addition to the 1993-94 equating process was the inclusion of performance events in arts & humanities and practical living/vocational studies. Because the initial year of testing for these content areas was 1992-93, performance events from 1992-93 and 1993-94 were administered as part of the equating study. Results from these content areas will be included in reporting for the second accountability cycle.

Slight changes were also made to the grade levels used in the 1993-94 equating study. For the 1992-93 equating students in grades 4, 8, and 12 were administered performance events from the corresponding grades. In 1993-94, grade 4 performance events were administered to grade 5 students. The rationale for this shift was that because the equating study was conducted early in the fall, grade 5 students would be more comparable than to the grade 4 students who completed the events in the spring. At grades 8 and 12, students from grades 7 and 11, respectively, were also included in the equating sample. This was done primarily to ensure that enough students would be available to complete each task. It was decided that at the middle and high school levels, the grade 8 and 12 performance events were appropriate for students in adjoining grades.

As a precaution against prior exposure to the performance events, schools used in the 1993-94 equating study were different from the schools used in 1992-93. Actually, the two studies were conducted in different states.

EQUATING SAMPLE AND ADMINISTRATION. The students participating in the equating study were drawn from a northeastern state. Volunteer schools were solicited with the assistance of local school officials and a school collaborative. All schools participated on a volunteer basis in return for training in the administration of performance testing and donation of the performance events materials. There were 26 elementary schools, 18 middle schools, and 14 high schools that volunteered for the study and whose staff participated in the training. Of these schools, 1 elementary school, 3 middle schools, and 5 high schools did not complete the testing.

The goal was to obtain a sample of 2,000 students at each of the three grade levels tested. With each student completing a task from the initial testing year (1991-92 or 1992-93) and the current year this would have produced a yield of 4000 completed tasks or more than 100 students per task. The goal of 2,000 students was not reached at any grade level. The number of students completing tasks within each content area by year is presented in Table 9-8.

TABLE 9-8
NUMBER OF TASKS COMPLETED BY GRADE, CONTENT AREA, AND YEAR

Subject Area	Grade 5		Grade 7/8		Grade 11/12	
	Initial Year	1993-94	Initial Year	1993-94	Initial Year	1993-94
Math	304	312	234	186	95	122
Science	280	396	181	276	95	129
Social Studies	260	408	298	372	106	131
Arts & Humanities	188	192	261	271	106	107
Practical Living/Voc. Studies	263	263	313	376	140	147
Total	1,295	1571*	1,287	1481*	542	636*

*Students completing performance events scored in two subject areas are counted twice; actual number of students is lower.

The table shows that the goal of 2000 students was not reached at any grade level. The most serious shortage was at grade 11/12 where only 542 students completed tasks from the initial year and 544 students completed 1993-94 tasks. Also evident from the table is that students did not complete two tasks from the same content area. The extreme differences in some cases between the number of students completing initial year and 1993-94 tasks raises additional questions about the equivalence of the groups completing the tasks from each year.

The equating study was conducted during the fall of 1994. Staff from schools participating in the study attended a training session in late summer to learn how to administer the performance events. Under conditions similar to the actual KIRIS assessment, teachers in the participating schools conducted the performance events testing. Student response materials were then scored by the same staff of trained scorers who participated in the scoring of performance events for the KIRIS assessment.

PROCEDURES. The following section describes and provides examples of the methods used in the equating of the 1993-94 performance events to the performance events from the initial year of testing. As will be described there are points in the process where subjective decisions are required in the assignment of performance level cut points. Given the extremely small size of the equating sample at grade 12, some of these decisions would have to be considered more arbitrary than desired. The equating procedures can be classified into five major categories which will be described in the following paragraphs.

1. Assign performance levels based on tasks from the initial year.
2. Compute frequencies of scores on 1993-94 tasks.
3. Convert frequencies to give equal weight to each task.
4. Assign performance level cutpoints to 1993-94 tasks.
5. Apply cutpoints to 1993-94 scores.

The equating sample contained no distinguished students in science at grade 8. This result is consistent with the results from KIRIS where only 0.04% of the students statewide were classified as distinguished in grade 8 science. This does, however, pose a problem

for equating purposes. The problem is solved by the convention of assigning the maximum possible score point on each task to the distinguished level (explained more fully at a later step). In all other content areas at least one student in the equating sample was assigned to each performance level.

Step 2: Compute frequencies of scores on 1993-94 tasks.

The second step in the process was to compute the score frequencies on each 1993-94 task within a content area. The score on a task was the sum of between one to six 4-point questions that comprised the task. Therefore, the individual performance tasks had score ranges which varied from 0-4 for single question tasks to 0-24 for 6 question tasks. At grade 4 all of the tasks were single question tasks. At grades 8 and 12 there was variation across and within content areas. The percentages for grade 4 math are presented in Table 9-10 as an example.

Step 1: Assign performance levels based on tasks from the initial year

The first step in the process is to assign students in the equating sample to performance levels based on their performance on tasks from the initial year. This provides the basis for the equating. The distribution of students across performance levels determined in this step will be used to determine cutpoints for the 1993-94 performance tasks.

Students were assigned to a performance level using the cutpoints established during the initial year of testing. These are the same cutpoints that were applied during the KIRIS assessment in the corresponding year. After all students completing a task have been assigned to a level, the percentage of students at each level across tasks within a content area was computed. The distributions of students across performance level by grade and content area are presented in

Table 9-9. As a point of information about the comparability of the equating sample students to Kentucky students, the distributions of students across performance levels on the appropriate KIRIS assessment are also presented.

TABLE 9-9
DISTRIBUTIONS OF STUDENTS ACROSS PERFORMANCE LEVEL BY GRADE AND CONTENT AREA

Grade	Content	Group	# EQ	Percentage of Students in Proficiency Category			
				N	A	P	D
4	Math	EQ	303	46.9	36.6	8.9	7.6
		KY		63.2	29.5	5.6	1.7
	Science	EQ	280	26.4	62.5	10.0	1.1
		KY		52.3	43.8	3.7	0.3
	Soc St	EQ	260	26.9	63.1	7.3	2.7
		KY		40.8	53.7	4.6	1.0
8	Math	EQ	234	40.2	38.5	15.8	5.6
		KY		62.2	23.9	10.6	3.3
	Science	EQ	181	48.1	50.3	1.7	0.0
		KY		62.9	36.1	0.9	0.0
	Soc St	EQ	298	28.9	58.7	12.1	0.3
		KY		42.2	48.5	9.0	0.3
12	Math	EQ	106	50.0	34.0	11.7	4.3
		KY		47.0	42.6	8.6	1.9
	Science	EQ	95	50.5	45.3	3.2	1.1
		KY		34.7	60.6	4.1	0.6
	Soc St	EQ	94	42.5	46.2	7.5	3.8
		KY		44.5	44.9	10.3	0.3

TABLE 9-10
PERCENTAGE, BY TASK, OF STUDENTS AT EACH SCORE POINT FOR GRADE 4 MATHEMATICS PERFORMANCE EVENTS

	# Students	Score Point				
		0	1	2	3	4
MA1	104	12.5	22.1	23.1	26.9	15.4
MA3	102	2.0	15.7	37.3	36.3	8.8
MA4	106	15.1	26.4	36.8	15.1	6.6

Step 3: Convert frequencies to give equal weight to each task.

In this step, the percentage of students at each score point across tasks within the content area was computed to be used in the assignment of cutpoints in Step 4. One way to do this would be to compute the percentage of the total math sample obtaining each score point on each task. That is, the number of students scoring 3 on task MA1 would be divided by the total number of students completing math tasks (312 at grade 4). The sum of these percentages across tasks would be 100%. The sum of the percentages within the task would be equal to the percentage of students in the total math sample completing that task. On task MA4 at grade 4, that would be 106/312 students. Using this method, tasks completed by more students would receive more weight. This would not be too much of a problem in grade 4 math where the students were fairly equally distributed among tasks. In other areas, however, the distribution of students in the equating sample was not quite as even.

To give equal weight to each task, the sum of the percentages within each task are made equal. That is, if there are four tasks within a content area, the sum of the percentages within each task is 25%. The sum of the score point percentages across tasks is still equal to 100%, but the sum of percentages within each task is equal. The adjusted percentages from the math tasks in Table 9-10 are presented in Table 9-11.

TABLE 9-11
PERCENTAGE, ACROSS TASKS, OF STUDENTS AT EACH SCORE POINT FOR GRADE 4 MATHEMATICS PERFORMANCE EVENTS

	# students	Percentage of students at each score point				
		0	1	2	3	4
MA1	104	4.2	7.4	7.7	9.0	5.1
MA3	102	0.7	5.2	12.4	12.1	2.9
MA4	106	5.0	8.8	12.3	5.0	2.2

Step 4: Assign performance level cutpoints to 1993-94 tasks.

The next step in the equating process was to use the percentages from Step 3 to assign performance level cutpoints to each of the 1993-94 tasks. The goal in setting the cutpoints was to match the distribution of performance levels determined in Step 1 based on the 1991-92 and 1992-93 tasks. Given the limited number of tasks and score points, it would be only by chance that one could hope to exactly match the distribution of percentages from the initial year. Also, given the matrix of score points and tasks, there are multiple patterns of cutpoints that one could use to arrive at a similar distribution of percentages. Although the process requires a degree of subjectivity, the following general guidelines were used in making assignment decisions:

- Keep the sum of Novice and Apprentice and the sum of Proficient and Distinguished as close as possible to the initial year.
- When forced to make a choice, place more students in the higher performance level rather than the lower level.
- Try to represent equally each performance level in each task.

The distributions of the equating sample across performance levels based on their performance on the 1993-94 tasks are presented in Table 9-12. The original distributions derived from the tasks from the initial year are also presented for comparison. This table shows clearly some of the difficulties in establishing cutpoints based on the equating sample. At grade 8, in particular, there were several performance levels which were obtained by few students in the equating sample. Examination of the KIRIS scores revealed that these score points were also rarely achieved this year by Kentucky students. The cutpoints assigned performance levels to each task are presented in Appendix 9-3.

Step 5: Apply cutpoints to KIRIS assessment

The final step in the process was to apply the cutpoints to the 1993-94 KIRIS assessment and assign Kentucky students to performance levels. Each student was assigned to a level based on their performance on the particular performance task completed.

Before those assignments were made, a final adjustment was made to the cutpoints determined from the equating study. As was done with the open-response tests, any student achieving the highest possible score on their performance task was assigned to the distinguished level. It was only necessary to apply this rule to one math task and one science task at grade 4. In all other cases, the maximum score would already be assigned to the distinguished level. The problem occurs primarily at grade 4 because of the limited number of score points used for each task. The statewide percentages of students at each performance level are presented in Table 9-13.

TABLE 9-12
PERFORMANCE LEVELS OBTAINED BY EQUATING SAMPLE STUDENTS FOR 1991-92 AND 1993-94 PERFORMANCE EVENTS

Grade	Content	Year	Number of Students	Percentage of Students at Performance Level			
				Novice	Apprentice	Proficient	Distinguished
4	Math	91-92	303	46.9	36.6	8.9	7.6
		93-94	312	43.7	41.1	10.1	5.1
	Science	91-92	280	26.4	62.5	10.0	1.1
		93-94	396	25.2	62.2	11.2	1.2
	Soc St	91-92	260	26.9	63.1	7.3	2.7
		93-94	408	22.9	67.0	7.0	3.1
8	Math	91-92	234	40.2	38.5	15.8	5.6
		93-94	186	41.3	33.6	20.3	4.9
	Science	91-92	181	48.1	50.3	1.7	0.0
		93-94	276	49.1	49.4	1.6	0.0
	Soc St	91-92	298	28.9	58.7	12.1	0.3
		93-94	372	25.2	60.4	14.2	0.3
12	Math	91-92	94	50.0	34.0	11.7	4.3
		93-94	122	49.4	34.3	10.1	6.3
	Science	91-92	95	50.5	45.3	3.2	1.1
		93-94	129	50.6	45.4	3.9	0.0
	Soc St	91-92	106	42.5	46.2	7.5	3.8
		93-94	131	42.6	47.4	9.0	1.0

TABLE 9-13
PERCENTAGE OF KENTUCKY STUDENTS AT EACH PERFORMANCE LEVEL,
1993-94 PERFORMANCE EVENTS

Grade	Subject	Novice	Apprentice	Proficient	Distinguished
4	Math	64.7	28.3	2.3	4.8
	Science	44.4	48.3	4.4	2.9
	Soc Studies	41.2	49.8	6.1	2.9
8	Math	42.1	29.6	19.3	9.0
	Science	48.7	49.0	2.1	0.3
	Soc Studies	27.2	51.5	20.2	1.1
12	Math	57.7	25.1	12.6	4.6
	Science	47.0	47.6	5.3	0.1
	Soc Studies	46.5	43.5	8.4	1.6

CHAPTER 10
REPORTING TO SCHOOLS AND DISTRICTS

In 1991-92, 1992-93 and 1993-94, preliminary results relevant to the first accountability cycle were released in September following the testing, with final results being released the subsequent winter. For 1993-94, the first release, in September 1994, provided results for the open-response tests as well as statewide (but not school-level) results for writing portfolios. The second set of reports, released in January 1995, included school-level results for writing portfolios, as well as results for performance events, and noncognitive data. These reports also combined the data into an Accountability Index for 1993-94. Because 1993-94 is the end of the first accountability cycle of KIRIS testing, schools' overall accountability scores were computed. These scores were compared to each school's improvement goal or threshold score to determine the school's level of success. Each school's accountability score for the 1992-94 biennium, as well as reward or assistance status, were included in the January reports. Reward checks were be distributed in April. As noted elsewhere, the assistance associated with failing to meet improvement goals in the first accountability cycle primarily entails additional report and planning requirements, in addition to additional resources to schools and districts.

District reports were issued at the same time as school reports, and differ from school reports only in terms of being based on all students in the district rather than all students in a school. However, districts with a single school at a given accountability grade will not necessarily have scores identical to the school at that accountability grade; the inclusion of scores from students who attend classes in a special learning environment are likely to change district results even in this case. The remainder of this chapter describes in detail the reports issued following the 1993-94 school year. A third set of reports, to be released later in 1995, will contain new baseline scores for the second accountability cycle.

SEPTEMBER REPORTS

Initial reports for the on-demand testing were prepared in September 1994. School, district, and statewide-level results were released to the news media, while results were reported to schools and parents at the student level as well. Materials included item-level reports for the multiple-choice items in each content area (reading, mathematics, science, and social studies), item-level reports for open-response tasks (open-response questions and on-demand writing), student listings, and parent letters. At the school and district levels, a single student assessment curriculum report was prepared describing performance in terms of the number and percentage of students at each performance level in reading, mathematics, science, social studies, and writing. These reports are all described in detail in the *Interpretive Guide*, published in September 1994, which accompanied the reports. A copy of the report is presented in Appendix 10-1.

Reports mailed in September included:

- Student listings
- Individual student report
- Summaries of cognitive results
- Curriculum reports
- A summary of student questionnaire information
- Subgroup analyses
- Item-level multiple-choice reports
- Item-level reports for open-response tasks

Reports contained information about all students enrolled or tested at a particular school. Results for students who were tested at a school, but for whom the school was not accountable, were shown on the printed reports but were not included in the computation of any aggregate information. These students were identified by an asterisk (*) printed next to their student identification number on each report. Students who were excluded from testing according to Department of Education policy were not included in any student reports. In a small number of cases, students were mistakenly identified as eligible for alternate portfolios and thus were excluded from September reports, although they had completed on-demand tests. Upon notification from District Assessment Coordinators, follow-up parent letters were sent for those students who had been mistakenly excluded. These students were included in the subsequent accountability reports.

Individual student reports were sent to schools to be forwarded to students' parents/guardians. These reports presented each student's performance level in each subject area relative to the statewide average for that grade. Schools employed a variety of methods to report individual student data. Some schools simply sent individual student reports to parents or guardians, some enclosed letters explaining results, and some asked parents/guardians to attend conferences, at which time results were explained in detail.

Cognitive result summaries presented the percentage of students within a school at each performance level for each open-response subject area. The number and percentage of non-tested students and students exempt from testing were also presented.

Curriculum reports presented the percentage of students at each performance level in reading, mathematics, science, and social studies, broken down by specific academic expectations.

All students in the accountability system were administered a questionnaire at the end of open-response testing. While the questions administered varied slightly from grade to grade, these questions included parent's educational level, types of courses currently being taken and total number of courses taken in each subject area, post-graduation plans, grades, homework, and participation in in-school activities. Overall results were reported in the September reports.

Subgroup reports presented performance level summaries broken down by relevant subgroups (male/female, white/non-white, Chapter 1/non-Chapter 1) as well as by selected responses from the student questionnaire.

On student listings, multiple-choice reports presented student responses to multiple-choice items in each subject area. A correct answer was noted with a plus sign ("+"), while an incorrect response was noted by listing which incorrect response the student chose. In addition, the percentage correct at the school, district, and state level was listed for each item.

Open response reports presented each student's score on each of the common and matrix items across all subject areas (including on-demand writing). Also, each student's performance level was listed for the common item scales in reading, math, science, and social studies . For each item, the mean score at the school, district, and state level was also listed.

Student reports presented a summary list of students along with each student's performance level (as calculated on common items only) in reading, mathematics, science, and social studies. Each student's Normal Curve Equivalent (NCE) on each subject was also presented. NCEs, which are calculated using both open-response and multiple-choice questions, are used only in Chapter 1 determinations. The number of students excluded from testing was also listed.

JANUARY ACCOUNTABILITY REPORT

In January 1995, the second set of accountability reports were released. The purpose of these two-page reports was to combine all KIRIS information about a school, compute the statistic by which schools were to be evaluated for the first accountability cycle, and inform schools of their success or failure in meeting their improvement goals. A copy of this report for a school is included in Appendix 10-2; district level reports do not differ appreciably.

In addition to presenting the accountability index, these reports indicated percentage Novice, Apprentice, Proficient and Distinguished at the school or district level by content area and type of testing (performance event, open-response test, portfolio). For open-response testing, this information repeated results originally reported in the September reports, only with appropriate adjustments made for students incorrectly included in or omitted from the earlier report. Because all performance events testing was matrix-sampled, and because no machine-readable, identifying student information was placed on performance event forms, results were reported at the school and district level, rather than at the student level.

To provide information by content area, a summary index for each cognitive area is presented, as calculated by multiplying percentage Novice, Apprentice, Proficient and Distinguished by 0, 40, 100 and 140, and summing the products. No further calculations are required to obtain a cognitive index for writing, as writing portfolios constitute the only measure of writing. Combining the results across performance events and open-response questions requires additional weighting: for reading, scores on common and matrix-sampled open-response items are weighted equally, 50% each; for mathematics, science and social studies, scores on common and matrix-sampled open-response items are also weighted equally, 40% each, with the remaining 20% being contributed by performance events.

As noted in the Interpretive Guide, results for the four noncognitive indicators used in the first accountability cycle are included as well: attendance, retention rate, dropout rate, and successful transition to adult life. (A fifth indicator, reduction of physical and mental barriers to learning, is currently under review for possible inclusion in the future.) As with the cognitive indices, procedures used for combining noncognitive data and approved by the Kentucky Board of Education yield a single Non-Cognitive Index. The average of six values --the five Cognitive Indices and the Non-Cognitive Index -- yield a school's Accountability Index for 1993-94. This number was averaged with a school's index for the 1992-93 school year to obtain the overall Accountability Index for the 1992-94 biennium. Each report also included a group of messages, tailored to the school's results, indicating the school's reward/assistance status. These messages are included in the *Supporting Technical Information for 1992-93 Accountability Reports* document.

CHAPTER 11
STATEWIDE SUMMARY OF RESULTS

Table 11-1 summarizes the final accountability index results for the first Accountability Cycle[1]. Tables 11-2 presents the same analysis for all schools by region, while Table 11-3 provides this information by content area, and table 11-4 presents results for non-cognitive indicators. The data in Tables 11-3 and 11-4 are based on weighted averages of school results. School achievement of rewards/assistance status is discussed in Chapter 15.

TABLE 11-1
STATEWIDE ACCOUNTABILITY INDEX SCORES

Grade Level	Baseline	1992-93 Index	1993-94 Index	1992-94 Average Index	Gain*	Percent Gain*
Grade 4	33.9	39.3	43.7	41.5	7.6	22
Schools with both Grades 8 and 12	33.9	38.3	43.6	40.9	7.0	21
Grade 8	37.4	39.9	44.9	42.4	5.0	13
Schools with both Grades 8 and 12	37.9	39.2	45.8	42.5	4.6	12
Grade 12	39.3	39.4	46.4	42.8	3.5	9
Total	36.4	39.4	44.8	42.1	5.7	16

* Gain and percent gain figures may differ from those calculated from reported Baseline and 1992-94 Index because they are based on data that was not rounded to one decimal place.

OVERALL ACADEMIC INDEX

Statewide performance showed impressive gains on the 140-point academic index scale. The average of the five subject area scores for Grade 4 rose from 21.0 to 32.9 during the first Accountability Cycle, a 11.9 point rise. Similarly, at Grade 8, the average index rose from 25.3 to 34.5, for a total gain of 8.8 points. At Grade 12, the overall rise was similar, with a 1993-94 index of 37.0, a 8.4 point gain over 1991-92. The largest gains are at grade 4, with progressively smaller gains in grades 8 and 12. This pattern is consistent with the proportion of post-KERA time students in each grade have spent in school. It is also consistent with anecdotal evidence regarding the rate of change within different levels of schools.

---

[1] Successful appeals may modify some of these data.

TABLE 11-2
STATEWIDE ACCOUNTABILITY INDEX SCORES, BY REGION

Region Area	Baseline	1992-93 Index	1993-94 Index	1992-94 Average Index	Gain*	Percent Gain*
Region One	36.7	39.8	44.6	42.2	5.6	15
Region Two	37.1	40.1	45.5	42.8	5.7	15
Region Three	36.4	38.3	43.8	41.1	4.6	13
Region Four	38.3	40.9	47.3	44.1	5.8	15
Region Five	38.2	41.3	47.3	44.3	6.0	16
Region Six	34.5	37.9	42.8	40.4	5.9	17
Region Seven	35.7	39.5	43.6	41.5	5.9	16
Region Eight	33.1	36.1	41.7	38.9	5.8	18
Statewide	36.4	39.4	44.8	42.1	5.7	16

*Gain and percent gain figures may differ from those calculated from reported Baseline and 1992-94 Index because they are based on data that was not rounded to one decimal place.

TABLE 11-3
ACCOUNTABILITY INDEX SCORES BY CONTENT AND YEAR

Grade Level	Year	Reading	Math	Science	Social St.	Writing	Non-Cog	Cog Only	Overall
Grade 4	1991-92	21.9	18.1	20.5	28.6	15.9	96.4	21.0	33.6
	1992-93	32.9	20.8	22.9	29.6	31.3	96.4	27.5	39.0
	1993-94	39.4	28.5	26.0	32.9	37.5	96.5	32.9	43.5
Grade 8	1991-92	36.9	23.8	15.6	30.0	20.2	96.7	25.3	37.2
	1992-93	40.0	29.3	17.7	26.7	28.3	97.0	28.4	39.8
	1993-94	46.7	35.2	20.9	38.3	31.3	96.8	34.5	44.9
Grade 12	1991-92	29.0	28	29.4	29.1	27.7	93.0	28.6	39.4
	1992-93	23.5	27.8	26.2	23.3	41.6	94.0	28.5	39.4
	1993-94	38.8	37.9	33.0	35.8	39.4	94.4	37.0	46.6

TABLE 11-4
STATEWIDE NON-COGNITIVE INDICATORS

Grade Level	Year	Attendance Rate	Retention Rate	Dropout Rate	Transition to Adult Life	Overall Non-cognitive Index
Grade 4	1991-92	95.8	0.8	n/a	n/a	96.4
	1992-93	95.6	0.6	n/a	n/a	96.4
	1993-94	95.8	0.6	n/a	n/a	96.5
Grade 8	1991-92	94.7	2.4	1.1	n/a	96.7
	1992-93	94.7	1.8	0.7	n/a	97.0
	1993-94	94.7	2.2	0.8	n/a	96.8
Grade 12	1991-92	93.8	5.1	4.5	89.6	93.0
	1992-93	93.6	4.9	4.6	92.6	94.0
	1993-94	93.3	5.2	4.7	94.1	94.4

CHAPTER 12
SPECIAL WRITING PORTFOLIO ISSUES

THE PLACE OF PORTFOLIO ASSESSMENT IN KIRIS

Portfolios occupy a key place in KIRIS, both as a means of assessment that directly taps student work in classrooms, and as a means for supporting educational improvement in classrooms, schools, and districts. KIRIS fully included writing portfolios in the first accountability cycle; mathematics portfolios and alternate portfolios were administered during the biennium to establish baselines for the next accountability cycle. Alternate portfolios are the assessment vehicle for students with what in general terms can be described as moderate to severe disabilities that prevent them from participating in regular classroom instruction even after all possible assistance and instructional adaptations have been provided. Such circumstances likewise prevent student participation in the regular KIRIS testing program.

Portfolios are closely integrated with local curriculum and instruction, since the contents of the portfolios arise from students' classroom work. In concept, portfolios are developed by students over long periods of time--months and perhaps years--with considerable support and direction by teachers. Portfolios, then, may reasonably be viewed as the "best work" developed by students in conjunction with their teachers, peers, and other resources.

In addition to providing assessment information, portfolios provide a powerful means to support teachers' professional development and school improvement. KIRIS portfolios are scored locally using a scoring guide that requires considerable content knowledge. This expert judgment-based, local scoring provides a wholly different model of teacher responsibility and involvement in accountability systems than centrally scored or multiple-choice tests. In many respects portfolios are the portion of the KIRIS assessment that most directly and comprehensively supports educational reform due to the strong connection to students' classroom experiences and strong involvement of teachers.

OVERVIEW OF ISSUES

Three main issues regarding portfolios have arisen during the first accountability cycle:

1. Establishing an operational program, including setting standards, specifying content, developing assessment instruments and procedures, supporting teacher understanding of the portfolio requirements, and administering the assessment program for over 140,000 students each year. These aspects of the writing portfolio program are described in previous chapters.

2. Increasing acceptance, use, and instructional value of the portfolios to students, teachers, and others. These issues are addressed to a limited extent in this chapter and in Chapter 15, *Consequential Validity*.

3. Monitoring and improving scoring accuracy and reliability. This is the main focus of this chapter. This chapter is limited to writing portfolios, because the activities of mathematics and alternate portfolios have-been focused on establishing baselines for the second Accountability Cycle. Mathematics and alternate portfolios will be reported in the next Technical Manual.

CHAPTER 12
SPECIAL WRITING PORTFOLIO ISSUES

THE PLACE OF PORTFOLIO ASSESSMENT IN KIRIS

Portfolios occupy a key place in KIRIS, both as a means of assessment that directly taps student work in classrooms, and as a means for supporting educational improvement in classrooms, schools, and districts. KIRIS fully included writing portfolios in the first accountability cycle; mathematics portfolios and alternate portfolios were administered during the biennium to establish baselines for the next accountability cycle. Alternate portfolios are the assessment vehicle for students with what in general terms can be described as moderate to severe disabilities that prevent them from participating in regular classroom instruction even after all possible assistance and instructional adaptations have been provided. Such circumstances likewise prevent student participation in the regular KIRIS testing program.

Portfolios are closely integrated with local curriculum and instruction, since the contents of the portfolios arise from students' classroom work. In concept, portfolios are developed by students over long periods of time--months and perhaps years--with considerable support and direction by teachers. Portfolios, then, may reasonably be viewed as the "best work" developed by students in conjunction with their teachers, peers, and other resources.

In addition to providing assessment information, portfolios provide a powerful means to support teachers' professional development and school improvement. KIRIS portfolios are scored locally using a scoring guide that requires considerable content knowledge. This expert judgment-based, local scoring provides a wholly different model of teacher responsibility and involvement in accountability systems than centrally scored or multiple-choice tests. In many respects portfolios are the portion of the KIRIS assessment that most directly and comprehensively supports educational reform due to the strong connection to students' classroom experiences and strong involvement of teachers.

OVERVIEW OF ISSUES

Three main issues regarding portfolios have arisen during the first accountability cycle:

1. Establishing an operational program, including setting standards, specifying content, developing assessment instruments and procedures, supporting teacher understanding of the portfolio requirements, and administering the assessment program for over 140,000 students each year. These aspects of the writing portfolio program are described in previous chapters.

2. Increasing acceptance, use, and instructional value of the portfolios to students, teachers, and others. These issues are addressed to a limited extent in this chapter and in Chapter 15, *Consequential Validity*.

3. Monitoring and improving scoring accuracy and reliability. This is the main focus of this chapter. This chapter is limited to writing portfolios, because the activities of mathematics and alternate portfolios have-been focused on establishing baselines for the second Accountability Cycle. Mathematics and alternate portfolios will be reported in the next Technical Manual.

MAIN RESULTS

Seven writing portfolio studies conducted by KDE/ASME are reviewed in this chapter. Each study addresses the issue of scorer reliability. These studies are:

1. "Inconsistent Scorer Study" (1992-93)
2. "Random Sample of Schools" (1992-93)
3. "Audit" (1992-93)
4. "Audit Review" (1991-92, 1992-93)
5. "Self-Audit" (1991-92, 1992-93, 1993-94)
6. "Requested Audit" (1991-92, 1992-93, 1993-94)
7. "Scoring Analysis" (1993-94)

The dates in parentheses refer to the year the portfolios in the study were written. These studies have involved rescoring over 60,000 portfolios. The main results have been:

1. Many schools have made dramatic improvements in both student performance and scoring accuracy.

2. Estimates of state-wide scorer reliability, based on percent exact agreement between original scores assigned by local teachers and review scorers, has improved from approximately .50 to over .70 in the three years studied.

3. Review scorers regularly achieve inter-rater reliabilities of .80 or greater, based on exact agreement. Other quality control measures of agreement range from .70 to .90, but all have improved over time.

4. Auditing procedures implemented by KDE/ASME were accurate in identifying schools with plausible differences and in rescoring the audited schools' portfolios accurately.

5. KDE/ASME have made significant improvements in research and quality control procedures since 1991. These improvements have been made primarily on the basis of increased experience with the writing portfolio program as it has been available.

As various studies have been conducted over a number of years, KDE/ASME have collected a wide variety of data to inform understanding of the use of portfolios for assessment and instructional purposes. Most importantly, these findings have led to an increased understanding of how these two facets are inexorably intertwined. As KDE/ASME have continuously monitored and redefined training and scoring procedures in an attempt to increase scoring accuracy and demonstrate reliability, the impact of assessment on instruction has become apparent.

The writing portfolio was initially designed to alter writing instruction at the classroom level. In order to have a significant impact on classroom instruction, the design of the portfolio had to require teachers to drastically reshape their vision of the important aspects of writing. Traditional writing instruction had focused, for the most part, on spelling, punctuation, capitalization, and sentence structure. The intention of the writing portfolio design and scoring criteria was to shift the instructional focus from these facets of writing to others that had traditionally received less attention in classrooms, e.g., focused purposes for writing, addressing authentic purposes and audiences, substantial idea development with relevant supporting detail, effective organizational strategies and transitions. At the same time that these messages of change were being communicated to teachers across the state, they were also in the process of being implemented in the first round of portfolio assessment.

As the assessment system was being put in place, it was generally understood that the over-riding purpose of the writing portfolio was to collect information about student performance using a more valid vehicle. The cumulative results of numerous studies demonstrates that, not only is the portfolio a valid picture of student performance, but it also provides invaluable information about instructional practice. Recent studies indicate that teacher understanding of appropriate instructional practice has a direct link to consistent application of standards in accountability scoring.

The 1993-94 Scoring Analysis demonstrated a high correlation between scorer accuracy rates and evidence of instructional practice, e.g., schools which demonstrated high accuracy rates also produced high student performance, schools which demonstrated low accuracy rates produced lower student performance. In addition, instructional analysis, through examination of student portfolios, demonstrated that portfolios portraying a clear understanding of the instructional shifts noted above came from schools which achieved the highest accuracy rates and most improved student performance of all schools studied.

The data support links between the instruction, performance, and scoring accuracy. Those teachers who understand the qualities of good writing tend to instruct accordingly and help their students become better writers. When teachers understand the qualities of appropriate instruction, they also tend to recognize quality writing during scoring and are better able to apply standards consistently, thus scoring more accurately. If a teacher can clearly recognize a proficient portfolio, instructional strategies to teach students to produce written work that meets these expectations are more easily identified and applied by the teacher.

Most notably, the results of the Writing Portfolio Audit and the Scoring Analysis for audit schools demonstrate that improvement in student performance and scoring accuracy can occur quickly, but must occur simultaneously for either to have substantial growth. Results demonstrate that audit schools, among the least accurate in terms of scoring during one year, ranked among the most accurate during the following year. This was accomplished through a recognition of need and targeted professional development. These same schools also demonstrated substantially larger increases in student performance than the average school, again providing evidence of the positive impact this assessment tool can have on instructional practice.

First accountability cycle data demonstrate positive movement in both reliability and consequential validity, although this movement is uneven at diffeent grade levels. Many factors may contribute to the uneven growth of performance and accuracy. Issues of pre-service training, content integration, understanding of the need for new information, and the developmental flexibility of learners all contribute to these variances. The evidence collected in seven studies conducted over the first Accountability Cycle points to the power of classroom-based portfolio assessment in bringing about educational change. The evidence gathered this far suggests that reliability of scoring will continue to improve as teacher understanding of writing grows, and as reliability increases, so will the quality of the instruction offered to Kentucky students.

After more than three years of learning for teachers, students, and KDE/ASME, a series of operating procedures has been developed to inform communications, training, and decision-making surrounding the use of portfolios in instruction and assessment.[1] These principles are as follows:

1) Portfolios play a central role in support of reform, as initially thought. Not only are they a means of collecting valid data, they are an integral part of bringing about positive, instructional change. They can sometimes provide teachers with information about their own understanding of writing and instruction.

2) Student work and teacher scoring mirror instruction.

3) Training is more effective with feedback. When teachers find out that they do not score accurately, they are more receptive to being trained.

4) Test results will not change substantially until teachers have a clear, deep understanding of what they are to teach.

5) Scoring is not difficult, but many with preconceived ideas about standards are difficult to train.

6) Feedback on scoring accuracy is important before holding schools accountable for their accuracy.

7) Many teachers do not change slowly over time. Often a paradigm shift occurs where the teacher moves quickly from one mode of thought to another. After this shift is made, the work of making changes in instructional practice and application of standards in assessment makes more sense and becomes more self-directed.

The remainder of this chapter on Special Writing Portfolio Issues outlines, in chronological order, the activities and data that have led to the current understanding of the essential balance between reliable and valid assessment of student performance.

INCONSISTENT SCORER STUDY (1992-93)

During the summer of 1993, a selected group of Kentucky teachers convened to rescore writing portfolios that had been originally scored by teachers throughout the state. Two primary activities took place; each was designed to maximize the quality of the scoring of writing portfolios that had been produced by students in grades 4, 8 and 12 throughout Kentucky. The first activity was an examination of the scores from teachers whose scores were inconsistent with (most different from) the rest of their district's scores. The second, described in the following section, involved the selection of portfolios from a random sample of schools from throughout the state, and an examination of the differences between these schools and a criterion group (the summer scorers).

Two other documents provide background to this process; the details in those documents are not repeated in this report. The procedures for district rescoring and the selection of inconsistent portfolios are described in *The Writing Portfolio Development and Scoring Process*, published by the Kentucky Department of Education in March, 1993. Details about the procedures used to select, train, and qualify scorers for the 1993 summer rescoring are provided in the *Writing Portfolio Summer Rescoring 1992-93 Final Report*.

---

[1] These operating procedures have been adapted from a list of operating procedures included in, "Reconceptualizing the Role of Portfolios in Implementing Reform," a presentation conducted by ASME at the CCSSO Conference on Large Scale Assessment, 1995.

IDENTIFICATION AND RESCORING OF WRITING PORTFOLIOS FROM INCONSISTENT SCORERS

During the 1992-93 assessment year, District Assessment Coordinators identified a random sample of portfolios from each teacher to be rescored. District Assessment Coordinators then shuffled the portfolios and delivered to each scoring teacher a sample of portfolios from other teachers within their districts. Teachers rescored these portfolios. The rescoring was completely blind; teachers did not know which students' portfolios they were rescoring, or the original scores the portfolios received. District Assessment Coordinators then delivered the rescoring forms of all teachers within the district to Advanced Systems.

Advanced Systems calculated consistency rates of original scores to rescores. Each teacher's scores were reviewed to determine the extent to which his/her scores on the rescored portfolios matched those of other teachers in the district. Advanced Systems delivered a list of those teachers whose scores were inconsistent with the other teachers in the district to District Assessment Coordinators. Of the 6,744 scoring teachers from 1,306 schools statewide, 784 were deemed inconsistent. District Assessment Coordinators informed those teachers whose scores were significantly discrepant from the other teachers within their districts. Those inconsistent scorers sent all of the student portfolios they scored to the District Assessment Coordinators, who in turn shipped those copies to Advanced Systems.

After logging in the portfolios, Advanced Systems delivered them to KDE. KDE convened Kentucky teachers who had demonstrated skill in scoring in the past. These teachers were required to complete an intensive training session and qualify using Consensus Portfolio. Consensus Portfolios had been pre-scored by the Writing Advisory Committee, the standard-setting group. The training and qualifying process and results are described more fully in the section on the "Random Sample of Schools."

After training and qualifying was completed, these teachers rescored the portfolios of all inconsistent scorers. The revised scores for these portfolios were delivered to Advanced Systems. An analysis was completed to determine whether the inconsistent scorers were also inaccurate. The scores of portfolios originally assigned by scorers found to be inaccurate during rescoring were altered in the accountability system records to reflect final scores assigned rescoring; the scores of scorers found to be sufficiently accurate during rescoring were not changed, so that accountability system records reflected those scores originally assigned by the teacher. Results of the initial portfolio scoring, with necessary revisions (based on rescoring results), were merged with other assessment data and reported back to the schools

Tables 12-1 through 12-3 provide the results of the rescoring activity. Each row provides the data for portfolios that were originally placed in a particular category, and each column shows the number of those portfolios that were rescored in each category. For example, the first row of Table 12-1 shows that 1,267 grade 4 portfolios were scored as a *Novice* by the original ("inconsistent") scorers. When those portfolios were rescored in the summer, 935 of them were rescored as *Novice*, 236 as *Apprentice*, etc.

TABLE 12-1
COMPARISON OF ORIGINAL SCORES TO RESCORES FOR INCONSISTENT TEACHERS GRADE 4

Original Score	Total	Summer Rescore					
		N	A	P	D	Blank	Inc.
Novice	1267	935	236	6	0	0	90
Apprentice	1584	715	736	62	1	0	70
Proficient	571	104	314	102	7	0	44
Distinguished	70	8	29	17	9	0	7
Blank	1	0	0	0	0	1	0
Incomplete	10	1	1	0	0	0	8
Total	3503	1763	1316	187	17	1	219

TABLE 12-2
COMPARISON OF ORIGINAL SCORES TO RESCORES FOR INCONSISTENT TEACHERS
GRADE 8

Original Score	Total	Summer Rescore					
		N	A	P	D	Blank	Inc.
Novice	1488	936	295	16	0	0	241
Apprentice	1474	578	621	110	10	0	155
Proficient	546	96	268	132	23	1	26
Distinguished	89	3	38	31	13	0	4
Blank	19	1	0	0	0	15	3
Incomplete	121	22	9	1	0	4	85
Total	3737	1636	1231	290	46	20	514

TABLE 12-3
COMPARISON OF ORIGINAL SCORES TO RESCORES FOR INCONSISTENT TEACHERS
GRADE 12

Original Score	Total	Summer Rescore					
		N	A	P	D	Blank	Inc.
Novice	736	482	129	17	0	0	108
Apprentice	1142	485	449	85	9	1	113
Proficient	569	127	224	153	22	0	43
Distinguished	91	5	21	48	14	0	3
Blank	4	0	0	0	0	4	0
Incomplete	44	12	1	0	0	0	31
Total	2586	1111	824	303	45	5	298

RANDOM SAMPLE OF SCHOOLS (1992-93)

To evaluate the effectiveness and accuracy of scoring statewide, two samples were drawn. The first was a sample of 200 students per grade whose portfolios were rescored by Advanced Systems scoring staff. The second was a sample of 40 schools.

Advanced Systems selected a random sample of 200 students per grade and delivered this list of students to District Assessment Coordinators. District Assessment Coordinators returned the selected portfolios to Advanced Systems. To ensure consistency with scoring in Kentucky, members of the Kentucky Writing Leadership Team conducted the training of the Advanced Systems scoring staff. Consensus Portfolios were used for all scorers to qualify. In addition, the remainder of the Consensus Portfolios were seeded throughout the scoring process to continuously monitor the scorers' application of Kentucky standards.

The second sample consisted of 40 schools per grade whose portfolios were rescored by Kentucky teachers under the supervision of KDE staff. Advanced Systems selected a random sample of 40 schools each for grades 4, 8, and 12. Advanced Systems requested the District Assessment Coordinators to send the portfolios for those schools. A maximum of 100 portfolios were requested per school; if more than 100 students in the school at the assigned grade level had completed portfolios, Advanced Systems selected a random sample of 100 students within the school. District Assessment Coordinators shipped selected portfolios to Advanced Systems. Advanced Systems logged in these portfolios and delivered them to KDE. These portfolios from the Random Sample of Schools Study were mixed with the portfolios from the inconsistent teachers and the two sets of portfolios were scored simultaneously. Of the 40 schools notified at each grade, portfolios were received from 39 schools at grade 4, 39 schools at grade 8, and 38 schools at grade 12. KDE convened Kentucky teachers and trained them to score. Consensus portfolios were used to monitor scoring consistency to the state standards. The portfolios were rescored and the scores sent to Advanced Systems for analysis. The information from the rescoring was used to evaluate the accuracy of statewide scoring.

TRAINING PROCESS

FIRST DAY TRAINING. Scorers were taken through a review of the Holistic Scoring Guide, benchmarks, rationale, and 1992-93 training materials by table leaders. In the training questions raised by scorers were addressed in a manner consistent with information disseminated in the past by KDE.

DAILY TRAINING. On subsequent days of scoring, all scorers were taken through a review of the Holistic Scoring Guide and benchmarks prior to beginning the day's scoring. This time was also used to address questions and concerns. Few questions were raised during this daily training, with the exception of the definition of incomplete portfolios.

QUALIFYING PROCESS. After the training session, each scorer completed a qualifying packet of 9 portfolios, the scores of which had been pre-determined by members of the Writing Advisory Committee. After reading and scoring each portfolio, scorers' were provided with rationale sheets which noted the pre-determined score and provided a detailed rationale for the score assigned. To qualify, scorers had to correctly score at least 6 of 9 portfolios, with a total of no more than 3 points/performance levels off, (i.e., scoring those portfolios incorrectly, and only off by one performance given on each portfolio). Scorers who scored 5 of 9 portfolios correctly or who were 4 points off were designated as "within 1" of qualifying. The results of the qualifying test are presented in Table 12-4.

TABLE 12-4
QUALIFYING PROCESS RESULTS

Grade Level	4	8	12
Qualified	35	18	15
Within 1	5	14	1
Not Qualified	3	8	3

Based on these results, the following steps were taken to assure qualification and accuracy of scorers.

- Those scorers who fell within the range of one were kept in the scoring pool, and table leaders began an immediate formal drift check of six portfolios per scorer. All of these scorers were found to have a high accuracy rate during the first formal check and ensuing checks. It may be that the stress of participating in a "testing" environment caused discrepancies. The scorers themselves stated that they "second-guessed" their first reaction to the portfolios, and "should have just stayed with their first score."

- Those scorers who did not successfully qualify participated in a formal retraining session immediately after the qualifying session with KDE and Advanced Systems personnel. These scorers then joined the pool, and table leaders immediately began a second qualifying round by checking the scores of the first six portfolios scored. The results of the second round of qualifying are presented in Table 12-5.

TABLE 12-5
RESULTS OF SECOND QUALIFYING ROUND

Grade Level	4	8	12
Average Accuracy*	84%	83%	94%
Qualified	7	20	4
Did Not Qualify	1	2	0

*Percentage of 6 formal check portfolios scored correctly after formal retraining.

- Three scorers did not qualify in the second round. They continued to score in the session, closely monitored and had continual formal 6-portfolio check throughout the entire scoring process. One of the three continued to have difficulty past the second full day of scoring. Table Leaders continued to individually train and monitor this score throughout the entire scoring period.

In addition, those people who qualified in the initial round and second round were monitored for scoring accuracy each day, by table leaders scoring at least a 20% sample of portfolio, randomly selected from each scorer.

SCORING PROCESS. On the second day table leaders held a brief review of the scoring guide and benchmarks. Scorers were instructed how to deal with paperwork, alerts, copy issues, and ID numbers. Scorers then scored portfolios individually and indicated their scores on rescoring forms. On occasions when they ran into portfolios they found difficult to score, they were asked to revisit those later in the day and, if necessary, turn them in for other scorers to score independently. The scoring rooms were quiet, and scorer behavior was serious, focused and professional. There were no lengthy group discussions or consensus scoring. Table 12-6 presents the number of portfolios scored at each grade level.

TABLE 12-6
NUMBER OF PORTFOLIOS SCORED IN RANDOM SAMPLE OF SCHOOLS

Grade Level	4	8	12
Total scored	6,896	7,014	6,783

Two reports were generated to assist districts in designing professional development that best meets the needs of scoring teachers. Each District Assessment Coordinator received a report identifying accurate and inaccurate scorers within the district, as well as individual teachers reports consisting of student ID numbers, initial scorers, and a cross-tabulation chart in similar to Table 12-7.

RESULTS. Tables 12-7 through 12-9 provide the same type of data presented in Tables 12-1 through 12-3. Tables 12-10 through 12-12 contrast the two sets of data. Those tables provide the percentages of portfolios that were scored at each performance level, given the original score, from both sources of rescored portfolios (the inconsistent teachers and the random sample). The data in Tables 12-10 through 12-12 show that the conditional probabilities for the two groups were very similar. While there was a higher degree of agreement between summer rescores and the original scores of the random sample than between summer rescores and the scores from "inconsistent" teachers, the difference between the original scorers of the random sample and the "inconsistent" scorers was not great. This result led to an examination of the hypothesis that inconsistency between scorers in portfolio scoring might be a strong indicator of inaccuracy.

The original purpose of conducting the comprehensive intra-district rescoring was to improve the overall quality of scoring. If a teacher scored significantly differently from the other teachers within the district, there was also a good likelihood that the teacher was scoring differently from all other teachers; therefore, the teacher's scores were not merely inconsistent with those of other teachers, but, more importantly, likely to be inaccurate when compared to the scoring standards. The data did not confirm this assumption. Inconsistent teachers were almost as likely to be accurate as any randomly-chosen teacher from around the state in the 40 school random sample. This result caused a re-evaluation of the portfolio scoring process. Intra-district rescoring was judged to be insufficiently informative to warrant the tremendous effort that went into it. As a result, along with a series of other changes designed to improve the portfolio scoring process, the Department announced in February, 1994, that intra-district rescoring would no longer be done.

TABLE 12-7
COMPARISON OF ORIGINAL SCORES FROM RANDOMLY SELECTED SCHOOLS TO RESCORES
GRADE 4

Original Score	Total	Summer Rescore					
		N	A	P	D	Blank	Inc.
Novice	644	528	103	1	0	0	12
Apprentice	886	426	408	40	0	0	12
Proficient	211	33	138	38	1	0	1
Distinguished	28	1	13	13	1	0	0
Blank	1	0	0	0	0	0	1
Incomplete	12	1	0	0	0	1	10
Total	1782	989	662	92	2	1	36

TABLE 12-8
COMPARISON OF ORIGINAL SCORES FROM RANDOMLY SELECTED SCHOOLS TO RESCORES
GRADE 8

Original Score	Total	Summer Rescore					
		N	A	P	D	Blank	Inc.
Novice	1032	745	110	3	0	4	170
Apprentice	878	364	375	54	0	0	85
Proficient	297	48	136	86	9	1	17
Distinguished	38	3	7	14	8	0	6
Blank	4	0	0	0	0	4	0
Incomplete	36	4	0	0	0	1	31
Total	2285	1164	628	157	17	10	309

TABLE 12-9
COMPARISON OF ORIGINAL SCORES FROM RANDOMLY SELECTED SCHOOLS TO RESCORES
GRADE 12

Original Score	Total	Summer Rescore					
		N	A	P	D	Blank	Inc.
Novice	825	589	94	4	1	0	137
Apprentice	1431	579	630	79	2	0	141
Proficient	772	92	376	223	13	0	68
Distinguished	105	0	28	51	17	0	9
Blank	24	0	0	0	0	23	1
Incomplete	61	9	15	2	0	3	32
Total	3218	1269	1143	359	33	26	388

TABLE 12-10
PERCENTAGES OF 1992-93 WRITING PORTFOLIOS RESCORED INTO VARIOUS CATEGORIES FOR INCONSISTENT TEACHERS AND THE STATEWIDE RANDOM SAMPLE
GRADE 4

Original Score	Group	Summer Rescore						
		N	A	P	D	Blank	Inc.	Total #
Novice	Inconsistent Teachers	74	19	0	0	0	7	1267
	Random Sample	82	16	0	0	0	2	644
Apprentice	Inconsistent Teachers	45	46	4	0	0	4	1584
	Random Sample	48	46	5	0	0	1	886
Proficient	Inconsistent Teachers	18	55	18	1	0	8	571
	Random Sample	16	65	18	0	0	0	211
Distinguished	Inconsistent Teachers	11	41	24	13	0	10	70
	Random Sample	4	46	46	4	0	0	28
Blank	Inconsistent Teachers	0	0	0	0	0	100	1
	Random Sample	0	0	0	0	0	100	1
Incomplete	Inconsistent Teachers	10	10	0	0	0	80	10
	Random Sample	8	0	0	0	8	83	12
Total	Inconsistent Teachers	50	38	5	0	0	6	3503
	Random Sample	55	37	5	0	0	2	1782

TABLE 12-11
PERCENTAGES OF 1992-93 WRITING PORTFOLIOS RESCORED INTO VARIOUS CATEGORIES FOR INCONSISTENT TEACHERS AND THE STATEWIDE RANDOM SAMPLE
GRADE 8

Original Score	Group	Summer Rescore						
		N	A	P	D	Blank	Inc.	Total #
Novice	Inconsistent Teachers	63	20	1	0	0	16	1488
	Random Sample	72	11	0	0	0	16	1032
Apprentice	Inconsistent Teachers	39	42	7	0	0	10	1474
	Random Sample	41	43	6	0	0	10	878
Proficient	Inconsistent Teachers	18	49	24	4	0	5	546
	Random Sample	16	46	29	3	0	6	297
Distinguished	Inconsistent Teachers	3	43	35	15	0	4	89
	Random Sample	8	18	37	21	0	16	38
Blank	Inconsistent-Teachers	1	0	0	0	79	16	19
	Random Sample	0	0	0	0	100	0	4
Incomplete	Inconsistent Teachers	18	7	0	0	3	70	121
	Random Sample	11	0	0	0	3	86	36
Total	Inconsistent Teachers	44	33	8	1	0	14	3737
	Random Sample	51	27	7	0	0	14	2285

TABLE 12-12
PERCENTAGES OF 1992-93 WRITING PORTFOLIOS RESCORED INTO VARIOUS CATEGORIES FOR INCONSISTENT TEACHERS AND THE STATEWIDE RANDOM SAMPLE
GRADE 12

Original Score	Group	Summer Rescore						
		N	A	P	D	Blank	Inc.	Total #
Novice	Inconsistent Teachers	65	18	2	0	0	15	736
	Random Sample	71	11	0	0	0	17	825
Apprentice	Inconsistent Teachers	42	39	7	0	0	10	1142
	Random Sample	40	44	6	0	0	10	1431
Proficient	Inconsistent Teachers	22	39	27	3	0	8	569
	Random Sample	12	49	29	1	0	9	772
Distinguished	Inconsistent Teachers	5	23	53	15	0	3	91
	Random Sample	0	27	49	15	0	9	105
Blank	Inconsistent Teachers	0	0	0	0	100	0	4
	Random Sample	0	0	0	0	96	4	24
Incomplete	Inconsistent Teachers	27	2	0	0	0	70	44
	Random Sample	15	25	3	0	5	52	61
Total	Inconsistent Teachers	43	32	12	2	0	12	2586
	Random Sample	39	36	11	1	1	12	3218

FURTHER ANALYSES OF DATA FROM THE RANDOM SAMPLE OF SCHOOLS

The random sample of schools was chosen to provide a range of information about the scoring of portfolios that could not be obtained through other means. Using the data to compare scores of inconsistent teachers with scores assigned by their peers, as described above, was one such application. The data from the random sample of schools also provided information, at the school level, about the accuracy and dependability of the information teachers were providing in 1993.

ESTIMATION OF STATEWIDE SCORES, 1991-92 AND 1992-93

An estimation of what the statewide scores would have been if all portfolios had been scored during the summer rescoring can be calculated through a process known as "recasting." To recast data, one applies the conditional probabilities obtained from the sample to the entire population of scores. Consider, for example, the testing results for grade 4. According to the results provided by teachers, a total of 49,567 Grade 4 portfolios were submitted: 1,033 students provided incomplete or missing portfolios, 19,251 were scored as Novice, 22,246 as Apprentice, 6374 as Proficient, and 663 as Distinguished. The information in Table 12-10 shows (with the portfolios rescored as "incomplete" excluded) that, of the 19,251 portfolios scored as Novice by the original teacher, 83 percent (or 15,978) would have been rescored as Novice and 17 percent (3273) would have been rescored as Apprentice, had the results from the summer study been extended. Similarly, the conditional probabilities obtained from Table 12-10 can be applied to all remaining performance levels. These calculations would result in the estimate that, upon rescoring, 28,965 of the portfolios would be judged as Novice, 18,170 as Apprentice, 2433 as Proficient, and none as Distinguished.

To simplify aggregation of data, Kentucky reports its results on an accountability scale that recodes Novice scores as 0, Apprentice as 40, Proficient as 100, and Distinguished as 140. Placed on this scale, the original scores for grade 4 would result in an index of 32.7. The index for the recast scores, an estimate of what the scores would have been for students statewide if all portfolios had been scored by Kentucky teachers during the summer session, was 19.6.

DIFFERENCES BETWEEN ORIGINAL SCORES AND RESCORES

Table 12-13 shows summary data for both 1991-92 and 1992-93, comparing the index of the scores originally assigned by teachers to the recast index from the summer rescoring. The results show that there were significant differences between the scores assigned by teachers statewide and the Kentucky teachers who participated in the summer rescoring. The results also showed that the gap between original scores and rescores grew considerably from 1991-92 to 1992-93. While there were real gains in portfolio scores, those gains were much less than those based on scores reported by teachers statewide.

The gap between reported improvement and actual improvement made it clear that additional work was needed to improve scoring accuracy and to provide a means of monitoring and correcting highly inaccurate school index scores to be used in the accountability system. Much of that additional work was directed toward the audit effort, which is discussed later in this chapter.

TABLE 12-13
STATEWIDE AVERAGES ON WRITING PORTFOLIOS

Statistic	Year	Grade		
		4	8	12
Portfolio index from scores as originally reported by teachers	1991-92	16.1	20.5	28.2
	1992-93	32.7	29.7	42.9
Portfolio index from scores after recasting using results of summer rescoring	1991-92	12.5	14.1	25.7
	1992-93	19.6	18.2	27.6
Reported improvement (Difference between 1991-92 and 1992-93 means of scores as reported by teachers)		16.6	9.2	14.7
Actual improvement (Difference between recast scores of two years)		7.1	4.1	1.9
Differences between 1991-92 and 1992-93 scores due to differences in scoring		9.5	5.1	12.8

In addition to determining the extent of these mean shifts, the data from the random sample of schools were used to explore other issues. One was the correlation between original scores and rescores. The correlation coefficients calculated from this rescoring process were .50 for grade 4, .54 for grade 8, and .55 for grade 12. A second statistic calculated was the correlation between mean original scores and mean rescores, using school as the unit of analysis. The correlations were .73 for grade 4, .59 for grade 8, and .47 for grade 12. These generally low correlations were an additional source of concern.

As another means of checking the accuracy of scores, some of the portfolios at the summer rescoring session were scored twice within the session itself. This provided the correlation between two sets of scores under conditions that were arguably more controlled than the scoring completed in the schools. Even under these more ideal conditions, the correlations between scores were quite low. For grades 4 and 8, the correlation was .51; for grade 12, it was .59. It is difficult to interpret the fact that these correlations were of about the same magnitude as the correlations between the original scores and the rescores.

CHARACTERISTICS OF PORTFOLIO DEVELOPMENT RELATED TO STUDENT PERFORMANCE

The research literature states that by scoring a significant number of portfolios in an intensive setting, teachers can see clear instructional strengths and needs portrayed within the student work. This was the case in this rescoring session. The scorers found a number of patterns consistent with other scoring sessions administered by KDE/ASME. Two reports were generated to assist districts in designing professional development that best meets the needs of scoring teachers. Each District Assessment Coordinator received a report identifying accurate and inaccurate scorers within the district, as well as individual teacher reports consisting of student ID numbers, initial scores and rescores, and a cross-tabulation chart similar to Table 12-7.

The following general categories of portfolios were identified by scoring teachers during the summer rescoring.

- Cookie Cutter Portfolios - These portfolios were a bulk of portfolios from the same classroom or school that looked virtually the same. All the entries focused on the same topic and/or prompt, and in many cases were close to a fill-in-the-blank format. These portfolios demonstrated minimal student investment and choice and were consistently lacking in proficiency.

- High Investment Portfolios - These portfolios were highly individualized and personal portfolios that demonstrated strong student investment and choice. They generally dealt with portfolio requirements in a unique and creative manner and were consistently of greater proficiency.

- Portfolio Week Portfolios - These portfolios were a bulk of portfolios from the same classroom or school that were clearly first draft pieces, and as noted by dated pages, were developed quickly and at the last possible date for completion. They demonstrated little time investment or completion of an appropriate writing process and were consistently lacking in proficiency.

- Content Deficient Portfolios - These portfolios may portray any of the previous patterns, be they positive or negative, but also contained content pieces that decreased the students' demonstration of proficiency. This was one of the most frequent problems seen in portfolios across all grade levels. It indicated that although effective instruction in writing is becoming more widespread, writing in study areas other than Language Arts is still consistently less proficient. It may also have indicated a lack of instructional integration and collaboration of Language Arts and content area to teachers within schools.

- Prompt Portfolios - These portfolios demonstrated a lack of student choice in generation of topics. They contained pieces that were based on teacher-generated prompts that were inappropriate and/or served to limit the students' ability to focus on a relevant purpose and/or audience. These types of portfolios may also have included a small number of pieces that demonstrated some choice and showed greater proficiency than those that rely on prompts. This was further evidence that student investment and choice increased the demonstration of proficiency.

AUDIT (1992-93)

Among the measures taken to assure consistent and accurate scoring of writing portfolios across Kentucky was a rescoring activity known as the "writing portfolio audit." The purposes of the audit were to determine whether existing feasible measures for identifying schools with potential scoring difficulties were effective, to bring identified schools into line as much as possible with Kentucky writing standards, and to reinforce the message that the accountability system requires and insists on rigorous application of scoring standards by the teachers who score portfolios. One purpose of the 1992-93 audit for writing portfolios was to explore data discrepancies for schools tested during 1992-93. Another purpose was to examine accuracy rates for accountability purposes. The process of rescoring/auditing also assists KDE in identifying areas of greatest need and increasing the accurate application of Kentucky standards.

RATIONALE AND CRITERIA. As described in *The Writing Portfolio Development and Scoring Process*, the strategy for state- level rescoring for 1992-93 was to focus resources on a limited number of schools, but to rescore large numbers of portfolios from each of those schools in order to have sufficient data with which to identify problems. As noted in that publication, the intent was to select schools to be audited by using the following criteria:

- portfolio scores substantially different from on-demand writing scores;
- portfolio scores substantially different from other content area scores;
- portfolio scores substantially different from previous year's portfolio scores.

When that publication was written, it was expected that those variables would have the greatest likelihood of identifying schools whose self-reported portfolio scores were most different from those resulting from an independent scoring. It was not until the data were collected from the summer rescoring, however, that it became clear just how strong those variables were.

During the fall of 1993, 105 schools were selected for full auditing of portfolio results. This selection process involved a calculation of cumulative difference based on three set criteria: 1) the difference between 1991-92 and 1992-93 Writing Portfolio results, 2) the difference between 1992-93 on-demand and portfolio results, and 3) the difference between 1992-93 content area and portfolio results. Fifty 4th grade, 25 eighth grade, and thirty 12th grade schools were selected due to having the greatest difference in results of the comparison formula. No other criteria were used to select schools.

COMPARING THE AUDIT SCHOOLS TO THE RANDOM SAMPLE. The first step in conducting the audit involved data from the summer rescoring of the random sample. As previously described, 40 schools per grade level were selected at random for that effort. After the rescoring was completed, the difference between the original scores and the rescores and the average difference for each school were calculated. The schools were then ranked based on the difference (from greatest to least) between the two sets of scores.

Next, a statistic was computed for each school in the state by aggregating the three criteria listed. The original plan was to audit 50 schools at grade 4, 30 at grade 8, and 25 at grade 12. Because the discrepancies between original scores and rescores, as determined from the summer rescoring, were greater for grade 12 than they were for grade 8, however, resources were reallocated to select 25 schools at grade 8 and 30 at grade 12. The schools at each grade with the highest values on this aggregated statistic were provisionally put on a list to be audited.

The immediate question that arose was whether any of the schools selected for the audit had, by chance, been selected for the summer rescoring. If so, it was crucial to know whether there were differences in the accuracy of scoring for audit schools when compared to the other schools in the state. If the differences were small or non-existent, other criteria would be necessary for selecting audit schools that more clearly identified real differences between those schools and others.

As it turned out, the differences based on the statistic originally chosen were dramatic. At grade 4, three of the 39 schools that sent portfolios for the rescoring sample had been selected for the audit. Those three schools appeared first, second, and fourth in the rankings of scoring differences between original scores and rescores. At grade 12, where there was the most incidence of discrepancy between original scores and rescores, the results were similar. Four of the schools from the grade 12 random sample had been identified for the audit. These schools placed first, second, third, and eighth in the ranking of scoring differences among 38 schools. At grade 8, which had significantly fewer differences between original scores and rescores than did grades 4 and 12, results followed the same direction but were not as extreme. In this case, only two of the audit schools were among the random sample; they were first and twenty-first on the list of 39 schools.

Clearly, the schools selected for the audit had significantly greater discrepancies in their scores than did others in the rescoring sample. As shown in Table 12-14, the mean differences for the audit schools were more than twice as great as the mean differences for other schools; the average mean difference for all grades was almost 30 points. These data led to the conclusion that the procedures for identifying schools with discrepant scores were effective, and the audit proceeded using the originally-proposed selection criteria.

TABLE 12-14
COMPARISON OF ORIGINAL AND SUMMER RESCORING RESULTS FOR AUDIT SCHOOLS AND OTHER SAMPLED SCHOOLS

Grade	Type of School	No. of Schools	No. of Portfolios	Mean Original Score	Mean Rescore	Mean Difference
4	Audited	3	114	60.9	26.5	34.4
	Not audited	36	1668	32.1	21.1	11.0
8	Audited	2	193	60.1	42.5	17.6
	Not audited	37	2092	28.0	19.6	8.4
12	Audited	4	323	63.2	31.7	31.5
	Not audited	34	2895	44.5	31.0	13.5

OBTAINING THE PORTFOLIOS. As noted above, the audit selection statistic was calculated for all schools in the state. Schools were ranked from high to low on this statistic, and the first 50 schools at grade 4, 25 schools at grade 8, and 30 schools at grade 12 were selected for the audit. Nine of those schools' portfolios had already been scored during the summer rescoring session. Notices were sent to the remaining 96 schools that they had been selected for audit, and they were requested to copy all their writing portfolios and deliver them to the assessment contractor, Advanced Systems. After a series of follow-up telephone calls, the portfolios were received from all 96 schools.

SELECTING AND TRAINING THE AUDIT SCORERS. Audit scoring was carried out in November, 1993, by Advanced Systems' scorers in New Hampshire. While a large number of people are trained to score for Advanced Systems, only a small portion of this group was invited to score the audit portfolios. Because of the great importance of this rescoring effort, this group was kept small enough to be tightly controlled. Experienced table leaders selected the scorers for the audit from among veteran scorers.

These scorers were the same scorers who scored the on-demand writing prompts and some of the open-response questions. Members of the scoring team have up to 6 years of professional scoring experience. Most of the scorers were former classroom teachers, and the others had been involved in other aspects of education.

The ultimate criteria for scoring came from the Kentucky Writing Advisory Committee. This group, consisting of writing educators from throughout the state, has guided every step of the writing portfolio process. In order to have solid exemplars against which other work could be evaluated, the Writing Advisory Committee was convened and asked to select portfolios that represented, by their unanimous agreement, each performance level. This group of portfolios, known as the "Consensus Portfolios," was a key element in ensuring that the scorers in New Hampshire scored accurately.

The New Hampshire scorers received the same training as the Kentucky scorers who completed the summer rescoring, using the same training materials used to train Kentucky teachers in 1992-93. To ensure consistency, members of the Kentucky Writing Leadership Team conducted the training. The scorers were required to complete a Qualifying Packet before they were allowed to score (the same used for Kentucky teachers to qualify to score during the 1992-93 summer rescoring session). Scorers were required to qualify for audit rescoring by achieving a high level of agreement (60% exact matches, no more than one performance level difference on remaining portfolios) with scores from a subset of the consensus portfolios.

In addition, a stratified random sample of 200 portfolios per grade level was collected and rescored by these same Advanced Systems scorers to determine the "Acceptable Scoring Range" for schools across the state. This information was used to make the accuracy determinations for those schools involved in the audit process. The remaining portfolios from the consensus portfolio set were also scored. These additional portfolios served as a control group and were used to maintain a check on scorers' adherence to Kentucky standards. Information about the number of consensus portfolios, by grade and proficiency level, and the number of times each portfolio was rescored during the scoring process is summarized in Table 12-15. For example, here are 12 Consensus portfolios in Grade 4. In the study they were scored a total of 46 times as part of the quality control process.

AUDIT PROCESS

The process for completing the audit entailed tables of 6 scorers, with a Table Leader, scoring school sets of portfolios. As scorers completed a small number of portfolios, Table Leaders rescored portfolios to check consistent application of Kentucky standards. When scorers completed one school's set of portfolios, those portfolios were removed and a new school set was begun.

ACCURACY OF NEW HAMPSHIRE AUDIT SCORING. Data from the scores assigned by the New Hampshire group to the control group of consensus portfolios provides information about the group's accuracy. Table 12-16 provides results of the scoring done by the New Hampshire scorers on those consensus portfolios. There was a high degree of agreement in scores between the New Hampshire scorers and the Writing Advisory Committee. Exact agreement achieved for all rescores was 83 percent at grade 4, 52 percent at grade 8, and 66 percent at grade 12. Agreement within one performance level was 98 percent at grade 4, 93 percent at grade 8, and 99.6 percent at grade 12. The correlation between consensus portfolio scores assigned by the New Hampshire scorers and the Writing Advisory Committee was .87 for grade 4, .57 for grade 8, and .79 for grade 12. These were far higher correlations than had been found, for example, between double scorings by Kentucky teachers during the 1993 summer rescoring session.

TABLE 12-15
CONSENSUS PORTFOLIO RESCORING ACTIVITY

Grade Level	Proficiency Level	Number of Portfolios	Number of Rescores
Grade 4	Novice	4	15
	Apprentice	4	16
	Proficient	3	11
	Distinguished	1	4
Grade 8	Novice	3	26
	Apprentice	3	29
	Proficient	3	26
	Distinguished	1	7
Grade 12	Novice	4	70
	Apprentice	5	87
	Proficient	3	54
	Distinguished	1	17

TABLE 12-16
CONSENSUS PORTFOLIO RESULTS

Grade Level	Proficiency Level	Novice	Apprentice	Proficient	Distinguished
Grade 4	Novice	15	0	0	0
	Apprentice	1	13	1	1
	Proficient	0	3	8	0
	Distinguished	0	0	2	2
Grade 8	Novice	19	6	1	0
	Apprentice	2	18	9	0
	Proficient	2	12	9	3
	Distinguished	0	3	4	0
Grade 12	Novice	53	17	0	0
	Apprentice	22	46	19	0
	Proficient	0	7	43	4
	Distinguished	0	1	8	8

In addition to examining the degree of correlation between New Hampshire scorers and the Kentucky Writing Advisory Committee, it also was important to determine whether the average scores that these two groups assigned to the portfolios were similar. That is, correlation statistics simply indicate whether the two groups would rank the portfolios in the same order — they do not provide information as to whether the two groups are following the same standards.

To make this determination, averages for the two groups were calculated. This was not as straightforward as it might seem, because a sample of portfolios is not representative of the state as a whole. To obtain more accurate estimates for the scoring of Proficient and Distinguished portfolios, it was necessary during these analyses, reflected in Table 12-16, to examine and rescore a disproportionately high percentage of portfolios at those levels. After recasting the data, it became clear that the Advanced Systems audit scorers very closely matched the scores of the Advisory Committee, although their scores were slightly higher than those of the committee. The differences in the indices were 0.9 at grade 4, 6.8 at grade 8, and 2.6 at grade 12. Table 12-17 shows the two indices.

TABLE 12-17
COMPARISON OF SCORING INDICES BASED ON CONSENSUS PORTFOLIOS

Grade Level	Advisory Committee Index	Audit Scorers Index
Grade 4	32.7	33.6
Grade 8	29.7	36.5
Grade 12	42.9	45.5

ACCURACY OF KENTUCKY TEACHER SCORING AND RESCORING, 1992-93

As mentioned earlier, a random sample of portfolios was rescored in Kentucky during the summer of 1993. That rescoring was done by Kentucky teachers specifically trained to participate in that process. The results showed that the portfolio score assigned by an average teacher in Kentucky was considerably higher than the score assigned by teachers participating in rescoring, who worked under controlled conditions. The differences between the two sets of scores was 13.1 points for grade 4, 11.5 points for grade 8, and 15.3 points for grade 12. This gap between the scores assigned by teachers and the rescores, especially since the differences varied greatly by school, was a source of considerable concern.

Unfortunately, no procedure was implemented during the summer rescoring session to directly link those scores to any external criterion group, particularly the Kentucky Writing Advisory Committee. So while it is clear that the summer scorers were giving the portfolios considerably lower scores than did teachers from throughout Kentucky, no evidence was collected to show that the differences indicated that the summer scorers were more accurate. However, a total of 584 portfolios rescored in the summer were also scored by the New Hampshire audit team. The two groups gave identical scores to 64% of those portfolios.

That lack of external criterion for the summer rescoring was corrected before the audit was done in New Hampshire by having the Writing Advisory Committee generate consensus scores on a number of portfolios. In addition, the same Advanced Systems' scorers who participated in the audit also scored a stratified random sample of portfolios that had originally been scored by Kentucky teachers. The sample consisted of 200 portfolios (50 at each performance level, as originally scored) at each grade level. The differences between the original scores and the New Hampshire scores were calculated. These differences were 19.4 points at grade 4, 15.8 points at grade 8, and 19.9 points at grade 12.

Those statistics provide two pieces of information. First, they provide evidence that while the Kentucky 1993 summer rescores were somewhat higher than the New Hampshire audit scores, the two scoring groups were fairly consistent in their standards. That is, at grade 4 the gap between the average teacher statewide and the summer rescorers (based on the summer random sample) was 13.1 points, and the gap at grade 4 between the average teacher statewide and the audit scorers (based on the stratified random sample) was 19.4 points. By using the average teacher statewide as a base, this reveals a difference of 6.3 points between the Kentucky summer rescorers and the audit team. Since the audit team matched up very closely to the Writing Advisory Committee, these figures provide evidence that, while the summer rescorers were somewhat easier in their scoring than the Committee, they were not far from that criterion group.

The graph that follows summarizes the differences in scores. It shows the relative averages of scores assigned by all participants in the Kentucky portfolio scoring process: the Writing Advisory Committee, the audit scorers, the Kentucky summer rescorers, and Kentucky teachers in both audited and non-audited schools. Note that the figure for the average Kentucky teacher is an actual figure, based on original scores of all writing portfolios scored in Kentucky. Figures for the other scoring groups are based on inference: they are the results of differences in scoring indices measured between score groups on specific samples of writing portfolios, recast to apply statewide. More specifically, using grade 4 figures:

- **Kentucky Writing Advisory Committee** — the consensus portfolios provided a comparison of the audit team's scores with the committee's scores; the difference in those scores was 0.9. If all portfolios had been rescored by the committee, we can infer its statewide average would have been 13.3 minus 0.9, or 12.4.

- **New Hampshire audit scorers** — recasting the stratified random sample shows the difference between the audit scorers and average Kentucky teachers as 19.4. If all portfolios had been rescored by the New Hampshire audit scorers, we can infer their statewide average would have been 32.7 minus 19.4, or 13.3.

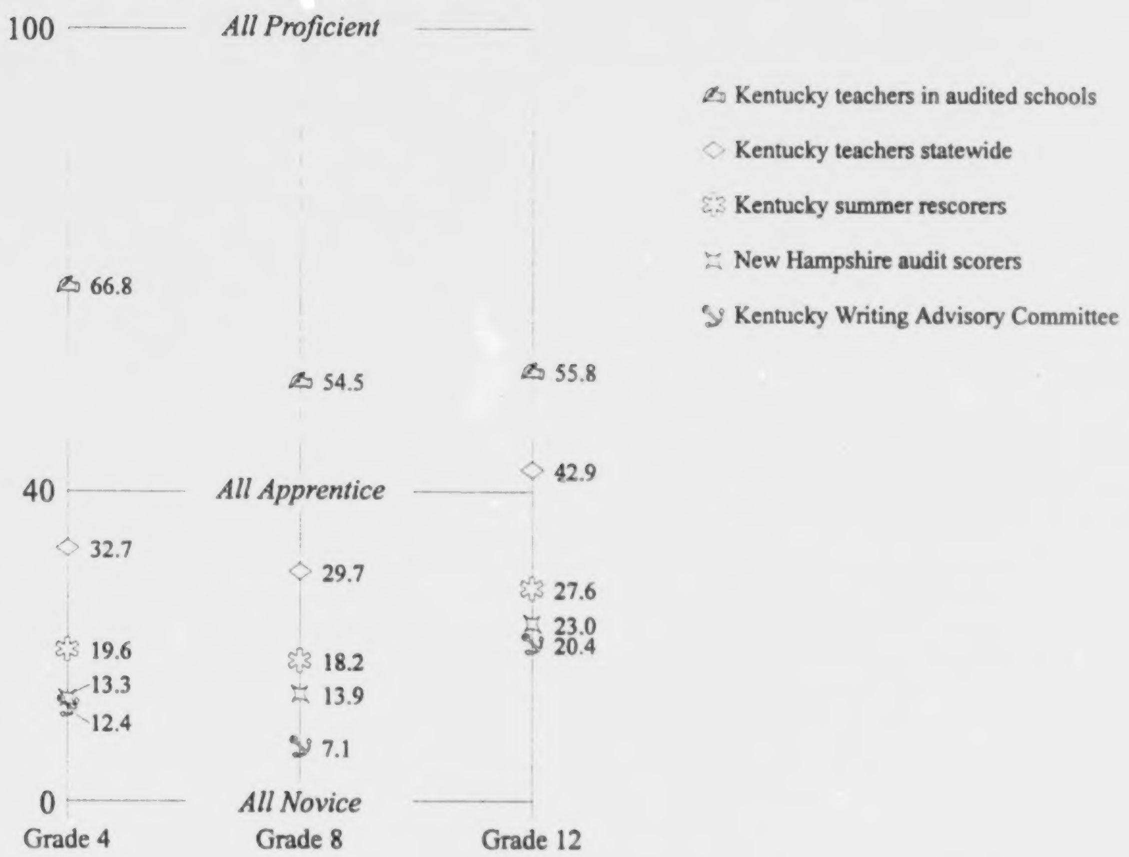
- **Kentucky summer rescorers** — based on the summer random sample rescoring, the difference between these rescorers and average Kentucky teachers was 13.1 points. If all portfolios had been rescored by the Kentucky summer rescorers, we can infer their statewide average would have been 32.7 minus 13.1, or 19.6.

- **Kentucky teachers statewide** — the index derived from actual portfolio scores was 32.7.

- **Kentucky teachers in audited schools** — the results of the audit scoring provided a comparison of the original scores with the audit team's scores; the difference in those scores at grade 4 was 53.5. If all portfolios had been scored by the teachers in the audited schools, we can infer their statewide average would have been 13.3 plus 53.5, or 66.8.

Placed on a scale ranging from 0 (representing all students assessed as Novice) to 100 (representing the Kentucky Department of Education's goal and expectation that all students be assessed as Proficient), Figure 12-1 shows dramatically that these different groups hold widely-varying opinions about how students performed in the portfolio assessment.

FIGURE 12-1
Average 1992-93 Kentucky Writing Portfolio Indices
Based on Scoring Standards as Implemented
by Different Sets of Scorers



DEFINING AN ACCEPTABLE RANGE OF DIFFERENCES BETWEEN ORIGINAL SCORES AND AUDIT SCORES

The differences between the audit scores and the average scores of teachers statewide were used to determine a range of acceptable differences in scores for each grade. That is, when differences between original teacher scores and audit scores were within that range, the original scores would be accepted by the Department of Education; when differences exceeded that range, the original scores would be considered too far in error to be accepted, and would be changed to the audit scores for accountability reporting.

Thus, in order for an audit school to have its original scores accepted, those original scores had to be no more different from the audit results than the average scores for the randomly-selected schools in the summer study. Using this standard, if the audit schools had a range of score accuracy typical of schools statewide, about half of them would have had their original scores accepted, and the other half would have had them changed. Of the 105 audited schools, however, 99 had differences outside the acceptable scoring range. For the other six schools, original scores were reported.

EFFECTIVENESS OF THE SELECTION CRITERIA

As can be seen from the results above, the effectiveness of the selection criteria was quite high. Almost all the schools that were chosen for auditing had scores that were far more different from the average group than would have been likely for a randomly-selected school. The results of the summer rescoring indicated that it was likely that the selection criteria were going to be highly effective; the results of the audit scoring confirmed that fact.

The selection criteria were so effective that results even correlated well within the audit group. That is, since the schools selected were all the most extreme in the state, there was much less variance among these schools than there would have been in a random statewide sample. But even for this limited group, the correlation between the selection statistic and the gap between the original scores and audit scores was .53 for grade 4, .74 for grade 8, and .27 for grade 12.

CONCLUSIONS AND ACTIONS

From a technical point of view, the audit worked extremely well. The training and qualification process for the audit scorers led to a team that was highly qualified, and was able to demonstrate their accuracy of scoring when compared to an essential criterion group (the Kentucky Writing Advisory Committee). The selection criteria obviously also worked very well, and identified schools whose scores most needed further examination.

One element that should be changed in the future is the process by which the portfolios were scored. For the purposes of this audit, a school's portfolios were all read at one time by the entire team of scorers. This allowed the scoring team to write and provide to the Department of Education a summary of observations for each school that might serve as useful feedback for the schools. As a result, scorers knew when they were scoring an audit portfolio and when they were, instead, scoring a consensus portfolio or a randomly-sampled portfolio. Consequently, the scoring process was not completely blind and can be criticized on this basis.

As a result of the audit, approximately 100 schools had their Writing Portfolio Index for 1992-93 adjusted. The results of the audit demonstrate several things:

- The criteria used to select the schools was effective in identifying those with the greatest difficulty in scoring. (This was not a random selection of schools.)
- Based on the rescoring of a sample of two hundred portfolios per grade level, we have identified a state-wide "Acceptable Scoring Range." This acceptable range refers to the level of scoring discrepancy we could expect from the average school at each grade level, based on 1992-93 scores.

- Some schools have difficulty in scoring. In many cases, this is a "one performance level too high" problem. This problem was addressed through the use of the High-End Portfolios included in 1993-94 training materials. It is this point that makes it crucial that Cluster Leaders stress the use of the High-Ends as scoring tools for all teachers. These portfolios should help alleviate the "one performance level too high" problem, because they show the upper limit of each performance level.

In addition, there were situations where portfolios were scored two or more performance levels too high. While new training materials could assist in rectifying this problem by including high-end portfolios and additional portfolio exemplars to assist teachers in gaining a further understanding of the language of the scoring guide, it is not reasonable to expect teachers to have completely put aside their own standards by the second year of a 20-year reform effort.

Scores were reported to schools in the form of full lists of student IDs, original scores, and audit scores. The schools also received a report on the 1992-93 Writing Cognitive Index originally reported from the school and as based on the audit results.

AUDIT REVIEW

Although the audit worked well to identify problems in scoring, some schools were extremely critical of the audit's results and questioned their accuracy. Consequently, schools whose writing portfolio scores were affected by the audit were given the opportunity to appeal their audit results. The Kentucky Department of Education designed a process for review of the 1992-93 Writing Portfolio Audit. In any high-stakes accountability system it is essential that there be a process for review. The Writing Portfolio Audit Review was designed to accomplish three main objectives:

- Determine the final writing portfolio scores for audited schools that chose to have portfolios rescored.
- Assess how consistent the audit scoring was with Kentucky standards.
- Assess how consistent statewide local scoring was with Kentucky standards for a sample of schools.

AUDIT REVIEW OPTIONS

In memos from the Department of Education distributed during February and March, 1994, audited schools were offered six options, listed in Table 12-18. For further details on the audit review process, refer to "Background Information: Writing Portfolio Audit Review Options," which accompanied the February memo.

All rescoring performed under options 2, 4, and 5 was done by Kentucky teachers using standards and benchmarks set by the Kentucky Writing Advisory Committee. If the option selected raised a school's accountability index for 1992-93, the school's baseline for the 1994-96 biennium of the second accountability cycle also rose.

Schools that chose options 1, 3, or 6 received their final accountability indices, including adjustment for option 3, in the March 10, 1994, release of scores. Scores for schools that chose options 2, 4, or 5 were presented in the January 1995 score report.

TABLE 12-18
WRITING PORTFOLIO AUDIT REVIEW OPTIONS

Option	Description	Number of Schools that Chose Option*
1	Accept the results of the writing portfolio fall audit.	4
2	Accept the results of the audit, but request that all 1991-92 portfolios be rescored. If an original score and the audit review score for that portfolio do not agree, the portfolio will be scored a third time. Scores for the 1991-92 portfolios will be the median of the three scores.	0
3	Request that the school's writing portfolio fall audit index be increased by the appropriate "acceptable scoring range" value (19.9 points for grade 12, 15.8 points for grade 8, and 19.4 points for grade 4).	28
4	Request that all 1992-93 writing portfolios be rescored where there was a discrepancy between the original score and the fall audit score, yielding an audit review score. For each individual portfolio there will be three scores – original score, audit score, and audit review score. If any two of these scores agree, that score will be considered valid and used in the writing portfolio index. If no two scores agree, the median score will be used.	1
5	Choose all procedures described in option 4, **plus** request that all 1991-92 portfolios be rescored. If original and audit review scores for a portfolio do not agree, the portfolio will be scored a third time. As in option 2, scores for 1991-92 portfolios will be the median of the three scores, with baselines adjusted as required.	11
6	Request that the school's original portfolio scores be restored.	54

* For six schools, the audit did not apply or did not change scores. One additional school entered into a separate agreement with the Department of Education, a controlled self-audit procedure.

RESULTS OF AUDIT REVIEW

At grades 4, 8 and 12, twelve schools chose options that resulted in a third scoring, by Kentucky teachers, for any portfolio for which the original score and the audit score disagreed. These schools delivered a total of 702 1992-93 portfolios for review scoring.

The review scoring team consisted of Kentucky teachers who were trained by Department of Education staff to adhere to the standards established by the Kentucky Writing Advisory Committee. After training and qualifying (by scoring preselected portfolios to the same standards as the Writing Advisory Committee), the review team began scoring. Included with the review portfolios from the audited schools were additional portfolios that had been pre-scored by the Writing Advisory Committee. The review team matched the Writing Advisory Committee's judgments very closely; 80 percent of the scores were exact matches. For the 20 percent of the scores that did not match exactly, 10 percent were too high, while 9 percent were too low. (Figures do not add up due to rounding error.) In summary, the audit review scorers exhibited a high degree of accuracy of scoring in agreement with the standards set by the Writing Advisory Committee; differences in scoring were not the result of a bias toward higher or lower scoring.

The scores produced by the audit review were substantially lower than the original scores. Approximately 90 percent of the scores were lowered by at least one performance level. On average across all portfolios, the scores differed by more than one performance level.

Table 12-19 summarizes the scores for the 702 disputed portfolios as assigned by the original teachers, the New Hampshire audit team, and the Kentucky teachers who served on the review team.

TABLE 12-19
WRITING PORTFOLIO PERFORMANCE LEVEL ASSIGNMENTS BY DIFFERENT SCORING GROUPS

Score Assigned	Group Doing the Scoring		
	Kentucky Teachers in Audit Schools	New Hampshire Audit Team	Review Team of Kentucky Teachers
Incomplete	6	2*	12
Novice	38	464	470
Apprentice	328	203	178
Proficient	268	33	38
Distinguished	62	0	4
Average Accountability Index	69.2	16.3	16.4

* Audit scorers were instructed to provide a score for every portfolio whenever possible, consequently scoring portfolios that were technically incomplete.

When viewed as an Accountability Index, the great similarity between the audit team and the Kentucky review team and the great disparity between the teachers in the twelve audit schools and the Kentucky review team are quite evident. The Kentucky review team assigned the disputed portfolios an index of 16.4. The index developed from the scores assigned by the New Hampshire audit team was almost identical — 16.3. The original scores assigned by the teachers in the audit schools were very different; theirs was an index of 69.2.

Table 12-20 shows numbers and percentages of exact agreements between the original scores and the audit review scores, and between the audit review scores and the scores assigned by the New Hampshire audit scorers. The percentage of exact agreement between the Kentucky audit review scorers and the New Hampshire audit scorers was high — as high as agreements achieved on rescoring studies within individual scoring groups.

TABLE 12-20
LEVELS OF AGREEMENT FOR AUDIT REVIEW SCORES WITH ORIGINAL SCORES AND WITH AUDIT SCORES

Number of Portfolios Rescored during Audit Review	Level of Exact Agreement: Original and Kentucky Audit Review Scores		Level of Exact Agreement: New Hampshire Audit and Kentucky Audit Review Scores	
	Number	Percentage	Number	Percentage
702	89	12.7	487	69.4

SELF-AUDIT AND REQUESTED AUDIT

BACKGROUND AND PURPOSE. Following the announcement of the Audit Review, KDE received some requests from schools that had not been audited to also have their writing portfolios reviewed. As a result, on May 13, 1994, KDE sent correspondence inviting all schools to indicate their interest in participating in a requested audit of their portfolios. Schools that indicated interest in requesting an audit were offered two options for the review process in a memorandum from KDE dated August 22, 1994. These "Request Review" schools chose either a centralized audit to be completed by ASME in Dover, N.H., or a self-audit to be completed by local school personnel under the monitoring and supervision of KDE staff. A highly successful pilot of the self-audit model had taken place in an elementary school in Northern Kentucky.

KDE held that there were strong instructional benefits in extending the self-audit to all schools. In both a self-audit and a centralized audit, portfolios are rescored to ensure consistency with the standards for writing portfolios set by the Kentucky Writing Advisory Committee. However, a self-audit also provides an opportunity for teachers to thoroughly understand the Kentucky Writing Portfolio standards and to improve their scoring accuracy. A self-audit session allows interaction among persons involved to address concerns expressed with developing and scoring writing portfolios. A key benefit is that teachers analyze the instructional implications for their own students, instructional practices, and school writing program.

SELF-AUDIT

PILOT STUDY RESULTS

The pilot study showed high success in every aspect for the writing portfolio self-audit model. The main results were:

1. Kentucky teachers can demonstrably score writing portfolios with very high consistency to the Kentucky standards established by the Kentucky Writing Advisory Committee under the conditions of this self-audit. The accuracy and acceptability of the scoring results generated through this model of self-audit can be endorsed without reservation by KDE. The teacher-scorers showed exact agreement over 80% of the time with the three indications of scoring quality instituted in this self-audit model:

- scorers agreed 86% of the time in double-blind scoring (i.e., 14% of the scores required conferencing between scorers who then agreed upon a score).

- the expert table leaders' second scoring agreed exactly 97% of the time with the scores assigned by the teachers. The table leaders checked 22% of total portfolios scored. The two table leaders were experienced teachers and scorers, and both members of the Writing Advisory Committee.
- exact agreement was 87% between the teachers and the quality control portfolios previously scored by the Kentucky Writing Advisory Committee. This level of agreement is slightly higher than that achieved by the New Hampshire audit team (84% for 4th grade) or the Audit Review scorers (approximately 80%). The scorers scored approximately one quality control portfolio for every five "live" portfolios.

2. The self-audit scoring process and results enjoyed very high acceptance and credibility with the school and district personnel, and were endorsed by the Kentucky Writing Program.

3. The self-audit process, combined with school and district efforts before the self-audit, produced teacher-scorers who were highly accurate and confident. Participants in the self-audit pilot have given numerous workshops and otherwise demonstrated enhanced leadership in writing portfolios, curriculum development, and writing instruction.

4. The detailed operational aspects of the pilot-test were effective, and could be easily extended to similar schools to provide acceptable quality controls, efficient administration, and valuable instructional feedback. The school termed the use of resources very cost-effective.

5. The Kentucky Writing Program estimated that it had sufficient staff resources to extend the self-audit model, as pilot-tested, to a significant number of schools, in addition to continuing its planned activities. The regional and state writing resource teachers have endorsed the concept.

6. The scores assigned in the self-audit largely agreed with the scores assigned through the Audit. The school's self-audit Writing Cognitive Index for 1992-93 was 11.1; the Audit assigned an index of 9.9.

7. Even though the self-audited writing portfolio results were lower than the school's original scores, the school showed considerable growth in student performance over three years. The school's Writing Cognitive Index increased over 500% from the baseline score of 5.9 in 1991-92, to 11.1 in 1992-93, to 31.9 in 1993-94.

These school index scores reflect a dramatic change in distribution of scores. The percent of scores fell from over 4 out of 5 students scoring Novice in 1991-92 to fewer than 1 in 3 by 1993-94. Concomitantly, the percentage of students scored Apprentice increase over four times, from 15% in 1991-92 to 61% in 1993-94. An indication of both the progress made by the school, and the remaining challenge, is that 1993-94 was the first year any Proficient portfolios were read (a total 8% of the students), and no Distinguished portfolios were identified.

IMPLEMENTATION IN A MIDDLE SCHOOL

To ensure that a self-audit could meet the highest standards of scoring quality, KDE/ASME designed a triple set of quality control scoring measures. These measures included double-blind scoring with moderation, extensive second reading by KDE staff, and scoring of portfolios previously assigned a score by Kentucky Writing Advisory Committee members. Altogether the self-audit required nine days. Two days were spent in training and qualifying. Seven days were spent on scoring and discussing instructional implications for the 1991-92, 1992-93, and 1993-94 portfolios.

TRAINING. KDE Writing Portfolio Consultants provided portfolio scoring training for the scoring team of six teachers from a middle school. Training progressed for most of the first day and followed the standard format using the Kentucky Writing Assessment Holistic Scoring Guide, 8th grade Benchmarks, Exemplar Portfolios, and High-End Portfolios. Trainers included a discussion of objectivity and bias issues.

QUALIFYING. Participants had to demonstrate they could accurately apply the Kentucky Writing Portfolio Holistic Scoring Guide in order to qualify to score in the audit. Scorers began qualifying in the afternoon of Day 1 and scored two qualifying portfolios. After each qualifying portfolio, KDE personnel lead a discussion of the rationales for the scores. Day 2 of the session began with recalibration to the scoring guide after which scorers completed the remaining seven portfolios of the Qualifying Packet. To qualify, scorers needed to correctly score 66%, or a minimum of six out of the nine Qualifying Packet portfolios with a total of no more than three points off on the entire packet. Only two out of six scorers qualified using the 9-portfolio packet. The remainder of the scorers qualified --i.e., reached 66% correct-- as KDE personnel "read behind" them during the actual scoring of portfolios to verify their scoring accuracy. It is important to note that the scorers who qualified this way were scoring portfolios that were nearly all Novice portfolios.

SCORING. KDE attempted to decrease the likelihood that teachers would score portfolios written by students in their own classes. The scheme used involved color codes. Each scorer was assigned a unique color. KDE staff attached a colored "flag" to each portfolio that indicated the original teacher who taught the student. (The flags were easily removable.) The first scorer attached another flag with his/her color, as did the second scorer. The arrangement of colored "flags" allowed the scorers and KDE personnel to quickly and easily control the flow of portfolios and balance the number of portfolios read by each combination of scorers.

Portfolios were scored in three, three-day sessions held two weeks apart. 1991-92 portfolios were scored in the first session, 1992-93 in the second, and 1993-94 in the third. The writing portfolio self-audit included the following elements for each of the three scoring sessions:

1. Portfolios were divided evenly among scorers in packs of 4, 5, or 6. The number of portfolios in a pack depended on available time and the total number of portfolios that had to be scored for a particular year. During most scoring sessions, scorers completed scoring two packs of portfolios per day, plus the quality control seed and the second scoring of portfolios from other scorers' packs. Altogether, each scorer usually scored approximately 22 portfolios each day; however, scorers never read more than 28 portfolios in one day.

2. Portfolios were scored twice using a "blind" procedure, where two scorers assigned a score independently of the other scorer. After both scores were assigned, the two scores were compared. Scorers engaged in "moderation", or a discussion session, if the two scores did not agree. The goal of moderation was for the two scorers to reconsider the standards and evidence and agree on a score. KDE personnel were called into these teacher-initiated sessions when another opinion was needed. The number of moderation sessions between teachers noticeably decreased during the second and third scoring sessions.

3. Quality control measures included scoring two consensus portfolios each day. The scorers' scores were compared with scores that had been previously assigned by the Kentucky Writing Advisory Committee. Scorer accuracy on consensus portfolios was 48%.

4. KDE personnel further maintained quality control by reading 60% of the portfolios scored. Virtually every portfolio with an "Apprentice" score or higher was read by a KDE staff member. Agreement between the scorers and the KDE staff on these "read behinds" was 90%.

5. The sessions included substantial time for discussion of instructional implications found in the portfolios. These discussions were led by KDE personnel.

Scorers discussed the problems found in portfolios by year, across years, as well as the possibilities for improving classroom instruction. Instructional improvements included ideas for managing writing instruction in content area classrooms, using working folders, and individual and school plans for future writing professional development.

OBSERVATIONS

The teachers and principals who participated in the Self-Audit were asked to reflect and respond to several questions that encompassed their expectations for the audit, insights they had gained during the self-audit, concerns about the writing program at their school, ideas for change, and ways to share with their colleagues.

The principal expected the teachers to share a common understanding of the scoring process, change individual and team attitudes, and immediately implement instructional strategies to improve writing and raise KIRIS scores. Although the scorers were clear as to the insights they had gained, the expectations were clearly the agenda of the principal. As the teachers shared what had become clear to them, they all agreed on several ideas:

- the writing portfolios showed some improvement over the three years
- the students need several years of instructional work to compile a strong portfolio
- the students need more freedom of choice in topic selection and ownership of the portfolio
- the writings needed to be more focused on a strong purpose and genuine audience
- the teachers need help to better understand the writing process and be able to teach their students to write clearly and effectively

Interestingly, the principal listed every change that collectively her scorers had recorded: a writing plan, student publications and recognition, writing stations and learning centers, portfolios at every grade level, more experiences for students, and common language across grade/content areas. An additional response from the principal indicates that the scorers are acting upon their insights, dreams and hopes. "The teachers who attended the self-audit have individually shared with me what they learned from the self-audit. They also brought me their charts [of possible actions] to type and share with the rest of the faculty. At our last faculty meeting team members shared their experiences and targeted areas we, as a school, need to improve upon. The team also has been proactive in getting charts made and providing assistance to content area teachers". Several observations were made by KDE personnel after three different weeks of working with the scorers:

1. Each school that participates in a self-audit will be a different entity. It cannot be assumed that the circumstances of schools that choose a self-audit will be similar. For example, even when a school requests a self-audit, it should not be assumed that all participants are equally committed to the process.

2. Reading and scoring portfolios is a powerful activity. Reading portfolios promotes insight that fosters an understanding of a school's writing program. Regardless of the attitude of a scorer when the reading begins, with time and supportive direction, the needs and strengths of instruction usually become evident.

3. Interaction and collaboration can build trust between teachers and KDE staff. The considerate, professional attitude of all staff involved can aid in building relationships that permit openness to generate and accept the benefits available from a self-audit.

4. Some difficulties arise when scorers are faced with reading primarily Novice portfolios. Reading portfolios from one's own school and scoring almost all of them Novice is demoralizing. It can also cause problems with scoring Quality Control portfolios accurately. This is particularly true when even Novice quality control portfolios are considerably better in quality than the portfolios being scored. When comparing the quality control statistics of this group to other scoring sessions, it is important to remember that for the Audit Review and Scoring Analysis studies, the scorers had considerably more experience and training than did the scorers at the self-audit school, and also had been exposed to many portfolios scored higher than Novice.

5. The true meaning of instructional implications becomes evident after reading three years of portfolios. After training, experience, and considerable collaboration, these scorers were able to write their own hopes, plans, and dreams for their writing program in their school. The power of instructional implications found in the portfolios came from the scorers' being able to put into words their own observations and to develop a group plan for accomplishing their goals.

These five observations support the idea that seeing, doing, and understanding supersedes the value of any mandated staff development. The professional development sessions these scorers now request or need will probably be perceived and accepted as necessary and welcome.

REQUESTED AUDIT

The Writing Portfolio Requested Audit was conducted to rescore and provide a revised Writing Cognitive Index for all schools requesting an audit. In addition, it was designed to provide participating schools with instructional feedback for 1993-94 portfolios.

The Requested Audit portfolios were scored by experienced writing scorers at Advanced Systems in Dover, New Hampshire. The scorers had experience in scoring Kentucky on-demand writing responses and Kentucky writing portfolios, as well as 5-9 years of experience in scoring student writing samples from other states.

TRAINING. Chief Readers, Table Leaders, and scorers were trained using the 1993-94 Reflection Packet and the Holistic Scoring Guide. Scoring training was conducted by Starr Lewis, Mary Alice Harmon, and Claudia Runge, from the Kentucky Department of Education. In addition, Table Leaders and Chief Readers were trained in operational and documentation procedures by Amy Buren from the Louisville, Kentucky office of Advanced Systems. All trainers monitored the first two days of scoring to ensure that all scoring and operational procedures were conducted according to the design.

QUALIFYING. Prior to beginning to score, all participants were required to satisfactorily complete a qualifying packet consisting of nine portfolios . Qualifying consisted of exact matches on six of nine portfolios, with no more than three performance level discrepancies on the complete qualifying packet. Those scorers who successfully completed the qualifying packet began scoring with a 20% rate of "back reading" by Table Leaders. Those scorers who did not successfully complete the qualifying packet participated in a review of the qualifying packet and began scoring with a 100% moderation rate. When non-qualifying scorers demonstrated an exact match on six consecutive portfolios, their moderation rate was be reduced to 20%, consistent with qualifying scorers. Qualifying rates are noted in table 12-21.

TABLE 12-21
QUALIFYING RATES

Grade Level	First-Round Qualifying	Monitored Qualifying	Total Scorers Qualified
4	71.5% (5 scorers)	28.5% (2 scorers)	100% (7 scorers)
8	83.3% (10 scorers)	16.7% (2 scorers)	100% (12 scorers)
12	75% (9 scorers)	25% (3 scorers)	100% (12 scorers)

PROCEDURE. Portfolios were spiraled in year-specific sets. This procedure reduced the possibility of any single scorer assigning scores to a disproportionate number of portfolios from any one school. They were then distributed in sets of 12, of which 11 of the portfolios were from audit schools, and one was a "seeded" portfolio, pre-scored by the Kentucky Writing Advisory committee, used to monitor consistent application of standards. Scorers scored two of these portfolio sets per day (approximately 24 portfolios). Scorers participated in daily activities to ensure calibration to the Kentucky standards.

Each year-specific set was completed prior to scoring portfolios from an ensuing year. As portfolios were scored initially, each rescore was checked to determine whether a second rescore was necessary. Any portfolio for which the rescore did not agree with the original score assigned by the school was scored again. The final scores assigned were determined using the following decision rule: any two scores that agree stand as the score of record; if all three scores differ the median score stands as the score of record. As year sets were completed, results written on scannable score forms were processed.

QUALITY CONTROL. Table Leaders conducted double scoring four times daily for each scorer. Each scoring was documented. If scoring discrepancies were noted, moderation occurred immediately, and moderation results were documented. Moderated scores were changed to reflect the score assigned through moderation by the scorer and the Table Leader or Chief Reader. The results of this moderation were used to verify the overall accuracy demonstrated by the scoring team.

Quality control 'Consensus' portfolios were seeded into the second and fourth portfolio sets twice daily. These portfolios are the same portfolios used for quality control in the 1992-93 Audit, the 1991-92/92-93 Audit Review, and the 1993-94 Scoring Analysis. Table Leaders checked scoring. If scoring discrepancies were noted, moderation (a discussion of the correct score and portfolio evidence) occurred immediately. The results of this scoring were used to verify the consistent application of standards across these varied activities. Results presented in Tables 12-22 verify the consistent application of standards within this activity, while data in Table 12-23 speak to consistency of this scoring with standards set by the Writing Advisory Committee.

TABLE 12-22
MODERATION RESULTS

Category	Grade 4	Grade 8	Grade 12
# of Portfolios Scored	165	342	325
Absolute Difference	.07	.11	.14
Difference	-.06	-.04	-.07
Exact Agreement between Table Leaders & Scorers	.93	.89	.86

TABLE 12-23
QUALITY CONTROL PORTFOLIO SCORING RESULTS

Category	Grade 4	Grade 8	Grade 12
# of Portfolios Scored	56	133	134
Absolute Difference	.38	.57	.34
Difference	-.09	+.02	0
Exact Agreement between Advisory Committee & Scorers	.62	.43	.66

REPORTING. Participating schools were provided with year-specific score reports including Student ID, Original Score, Rescore, and for 1993-94 portfolios, Instructional Analysis Annotations. An explanation of the training and scoring processes noted above was given as well.

WRITING PORTFOLIO SCORING ANALYSIS (1993-94)

The 1993-94 Writing Portfolio Scoring Analysis was designed to fulfill several purposes. First, schools involved in the regional audit meetings repeatedly had requested feedback about scoring prior to inclusion in any audit. In addition, these same schools had expressed a need for more substantial feedback from scoring sessions beyond score/rescore reports (e.g. scoring patterns, professional development recommendations, instructional recommendations). In addition to meeting these needs for a large sample of schools, the session also afforded the opportunity to provide teachers from across the state with extensive and appropriate professional development, and to inform and focus training plans for upcoming years. In an effort to strengthen various tiers of the training system, a large number of Cluster Leaders and Regional Coordinators were included in this session.

The 1993-94 Writing Portfolio Scoring Analysis took place in July, 1994, with a team of Kentucky teachers, who scored and analyzed a sample of Writing Portfolios from 50% of the schools in each district in Kentucky. Its results-- the product of a two-week scoring and analysis session involving 157 teachers, and KDE and Advanced Systems personnel-- provide a broad picture of state-wide scoring accuracy.

PARTICIPANT SELECTION. In the spring of 1994, Cluster Leaders received letters inviting them to participate in this session. Regional Coordinators received a letter asking them to serve as Table Leaders. A call for scorers appeared in the May issue of Kentucky Teacher. Table Leaders and scorers were selected from all trained teachers that applied. Many Regional Coordinators elected not to participate in the session, due to previous professional or personal commitments. Therefore, Table Leaders were selected from Cluster Leaders known to be experienced in this role, as well as those who had extensive experience with scoring and rescoring. The majority of scorers were Cluster Leaders, and additional scorers were recruited from prior scoring teams and local districts.

TABLE LEADER TRAINING. Table leaders attended a one-half day training session prior to the beginning of the general session. Scoring training included review of the 1992-93 Reflection Packet and a discussion of the Holistic Scoring Guide. Analysis training included practice scoring of full school sets (modeled by KDE and Advanced Systems personnel), guided completion of School-Level Reports, and practice in completing the required quality checks and related record keeping. Table Leaders were assigned to tables consisting of four to five scorers.

SCORER TRAINING. Scorers participated in a training session lasting approximately 3 hours, using the Holistic Scoring Guide, grade-specific Benchmarks, and the 1993-94 Reflection Packet. Much emphasis was placed on a discussion of the inter-related nature of scoring criteria and the processes for scoring "difficult" or "borderline" portfolios (i.e. High-ends). In addition, there was a thorough discussion of incomplete portfolios. Training was conducted by KDE and Advanced Systems personnel.

QUALIFYING. Scorers completed a qualifying packet of 9 portfolios, the scores of which had been pre-determined by members of the Writing Advisory Committee and/or KDE and Advanced Systems personnel. It was necessary to replace several qualifying portfolios that had been selected by the advisory committee as they were included in the Reflection Packet. During previous sessions the Reflection Packet was not available or used for training, so it had previously been unnecessary to replace them. Out of the 27 qualifying portfolios, 3 were selected by KDE and Advanced Systems personnel. This occurred because the writing Advisory Committee had met only once prior to this time to select portfolios, and, consequently, the portfolio pool was limited and used for multiple purposes (e.g. Reflection Packet portfolios, Qualifying Packet portfolios, and Quality Control portfolios).

Scorers were provided with rationale sheets which noted the pre-determined score and detailed the rationale for the assigned score. Scorers were expected to correctly score 6 out of 9 portfolios, with a total of no more than 3 points off on the entire packet. Scorers who did not qualify began scoring with their table and received continual monitoring by Table Leaders until their scores demonstrated sufficient accuracy. Results of the initial qualifying round are shown in Table 12-24. Scorers who did not initially qualify participated in a retraining session and began to score with a 100% read-behind by Table Leaders. All scorers qualified within the first round of general scoring.

MONITORING. Table leaders scored one portfolio per scorer for each school's packet (20% monitoring), as well as additional portfolios for any scorer who had not initially qualified, or if scorers requested an additional review. Table Leaders were able to accomplish this task well within the time allotted. Results of the portfolio monitoring are shown in Table 12-25.

TABLE 12-24
QUALIFYING RESULTS

Grade Level	Initial Qualifying Rate (%)
Grade 4	62%
Grade 8	70%
Grade 12	69%

TABLE 12-25
PORTFOLIO MONITORING RESULTS

Statistic	Grade 4	Grade 8	Grade 12
Number of Portfolios	1742	866	717
Absolute Difference	.10	.13	.14
Difference	+.01	+.04	+.04
Percent of Exact Agreement	.90	.87	.87

QUALITY CONTROL. Quality control portfolios with pre-assigned scores were read, scored, and discussed at each table twice daily, with the exception of 8th grade. At the 8th grade level, Quality Control Portfolios were scored once a day because 50% of the Quality Control Portfolios were a part of the Reflection Packet. (This was an unforeseen problem, and was addressed prior to any further scoring session.) This activity helped scorers continue to key in to Kentucky standards and reduce scorer drift. In addition, it provided data concerning the overall adherence to standards throughout the session, as shown in Table 12-26.

TABLE 12-26
QUALITY CONTROL RESULTS

Statistic	Grade 4	Grade 8	Grade 12
Number of Times Scored	626	170	305
Absolute Difference	.25	.33	.25
Difference	+.05	-.04	+.08
Percent of Exact Agreement	.75	.77	.77

DOUBLE SCORING. In addition to the quality control measures mentioned above, 42 school sets were double-scored and analyzed to check consistency of scoring and the reliability of the report form provided at the school level, and discussed below. This quality control measure was not initially included in the session design because it was unclear whether there would be time to complete this process. Thus, double-scoring was delayed until the end of the session to ensure that all school sets were analyzed at least once. Results of the double-scoring are shown in Table 12-27.

TABLE 12-27
DOUBLE-SCORE CONSISTENCY RESULTS

Statistic	Grade 4	Grade 8	Grade 12
Number of Portfolios Scored	480	660	1076
Absolute Difference	.14	.17	.18
Difference	+.01	+.03	+.01
Percent of Exact Agreement	.86	.83	.82

Not surprisingly, further analysis of the double-scored school reports demonstrated that, when scoring was consistent for individual portfolios, reported information about schools' instructional patterns was also consistent.

SCHOOL-LEVEL REPORTS. Participating schools were provided with two types of analyses, scoring and instructional. Each area was addressed by identifying patterns of strengths and needs, recommendations, and resources. All sections of the report were structured around the criteria of the Holistic Scoring Guide and the requirements of the writing portfolio and were intended to demonstrate to schools the dependent relationship between curriculum, instruction, and assessment.

Each day scorers at 8th and 12th grades scored and analyzed sample portfolios from three different schools, while scorers at the 4th grade level completed four school sets. When scorers completed all portfolios within the school, Table Leaders recorded the rescores on the report form. After each school set was rescored and scores recorded, Table Leaders facilitated a discussion centering on scoring and instructional issues. The discussion and report recommendations reflected impressions gained from scoring the sample, and the areas indicated on the report form were presented to schools as informed opinions, not mandates, or thorough pictures of scoring and instruction in the schools. The following areas were included in the recommendations sent to schools:

- Scoring Accuracy: Scorers were notified of the portfolios for which their scores differed from the original scores submitted by the school. Scorers were instructed to reexamine these portfolios to determine the criteria which most affected the score they assigned. Table Leaders did not complete the Scoring Accuracy segment of the report because of the technical nature of the process. This was completed by KDE & Advanced Systems personnel after the session was completed.

- Scoring Training Needs: Based on the reexamination of the aforementioned discrepant portfolios, tables identified the scoring criteria which had most often affected their decisions to alter scores. 81% of the schools in the sample misapplied the Purpose/Audience and/or Idea Development/Support criteria. When considering the increased scoring accuracy at the 4th grade level, and that 50% of the sample schools were 4th grade schools, this information is even more significant, in that even when scoring is more accurate, these criteria continue to create difficulties. Misapplication of these criteria suggests a general lack of understanding of authentic writing activities and what constitutes sufficient development of pieces.

- Instructional Strengths and Needs: After the scoring analysis was completed, Table Leaders facilitated a discussion focusing on the instructional evidence in the school set. Scorers identified instructional strengths and/or needs that were reflected consistently across the school set. Instructional strengths and needs generally were consistent with scoring strengths and needs. This confirms the necessity for continued professional development in the areas of Purpose/Audience & Idea Development Support , along with more focused scoring training.

- Instructional Recommendations: After all scoring and instructional analyses were completed, Table Leaders identified areas of instruction on which they recommended schools focus their instructional resources. These recommendations included steps in the writing process and types of writing (e.g., personal narrative, fiction, poetry).

- Scoring and Instructional Resources: The final segment of the report included recommendations for further review of print and video resources already available in local districts, as well as those which will be made available this year.

SCORING ACCURACY AND STUDENT PERFORMANCE. Scoring accuracy and student performance data, shown in Table 12-28, suggest that from 1992-93 to 1993-94, scoring accuracy increased significantly at Grade 4, and modestly at Grades 8 & 12. In addition, the data suggest that student performance has increased substantially at Grade 4, modestly at Grade 8, and has not increased at Grade 12.

TABLE 12-28
SCORING ACCURACY AND STUDENT PERFORMANCE

Statistics	Grade 4		Grade 8		Grade 12	
	1992-93	1993-94	1992-93	1993-94	1992-93	1993-94
State Average as Determined by Original Scores Submitted by Teachers	32.7	37.5	29.7	31.2	42.9	39.3
State Average as Determined by Recasting using Summer Rescoring	13.3	28.5	13.9	17.1	23.0	22.0
Difference between Original Scores and Recast Scores	19.4	9.0	15.8	14.1	19.9	17.3

OBSERVATIONS ABOUT TEACHER TRAINING

Data presented above are consistent with other data already available about differing grade levels. In many regards, a consistent relationship between the growth in student performance and scorers' abilities to score accurately is evident.

Consider the finding that particular scoring criteria that are consistently misapplied by inaccurate scorers (Purpose/Audience, Idea Development/Support), and that these same criteria are consistently listed as instructional needs on school-level reports. In reviewing the Grade 4 results, it appears that student products reflect improved instruction in terms of authentic purposes, audience awareness, and idea support. Paralleling this evidence of improved instruction is increasingly accurate scoring. In looking at the Grade 8 & 12 results, there is evidence of little instructional or scoring accuracy improvements. Indeed, the patterns of misapplication and instructional needs are consistent.

One conclusion to draw from these results is that teachers must understand the qualities of good writing before they can either instruct their students, or recognize these qualities in a scoring situation. Kentucky teachers' most significant area of need probably lies in instructional professional development. It cannot be assumed that altering or increasing scoring training, without addressing basic professional development needs in the area of instruction, will completely solve scoring accuracy problems or have a substantial effect on long-term instructional growth.

Although one may question the quality of training at this point, Grade 4 data suggest that training has facilitated significant improvements in both instruction and scoring accuracy. But why has this improvement not occurred to as great an extent at other grade levels? Perhaps the greatest change in instruction has taken place at the primary level, and consequently students may be more experienced with a varied learning environment and have increased writing skills due to early exposure to writing process. Grade 4 teachers may be altering their instruction more rapidly than secondary teachers as they receive students with different learning behaviors and performance strengths.

Perhaps another factor is the integrated nature of curriculum and instruction, and the self-contained classroom most often seen at the elementary level. Portfolio development requires inclusion of content area writing at all grade levels. Elementary teachers, in an effort to manage the full curriculum within a single classroom, have traditionally provided integrated instruction on a regular basis. This makes the content requirement of the portfolio a more natural part of instruction. Conversely, at the secondary levels, instruction continues to be compartmentalized, making inclusion of content pieces more difficult to achieve successfully.

In addition, while 8th and 12th grade Language Arts teachers have felt full responsibility for portfolio development, content area teachers may feel less responsibility than the content requirements of the portfolio would ideally assume. This may be especially true and specifically important at the 12th grade level since students are required to include 2 content area pieces. If these pieces demonstrate less proficiency (as supported by anecdotal findings from across the state), this third of the portfolio can have a detrimental effect of the overall score.

Perhaps another consideration is differing pre-service experiences for elementary and secondary teachers. Pre-service instruction for elementary teachers many times focuses on instructional strategies to be implemented regardless of the content focus, while pre-service instruction at the secondary levels often focuses on the mastery of content information, with less focus on essential instructional strategies.

Experience in the field suggests that elementary teachers are more receptive to new instructional methods and alternative forms of communication other than traditional essays and research reports, and more readily integrate new training into their classroom and scoring practices. Secondary teachers, most specifically high-school teachers, may have a much narrower view of what constitutes "good" writing and appropriate writing instruction.

With all of these points in mind, a logical strategy for improving the educational use of portfolios is to target delivery of instructional professional development at the secondary level. Continuing delivery of large-scale professional development activities that present a combined focus on curriculum, instruction, and assessment is indicated as well. Increasing interaction with and training for secondary teacher educators falls is consistent with this approach.

DISCUSSION OF TECHNICAL ISSUES OF SCORING QUALITY

Changes in writing portfolio scores -- differences between original scores as submitted by the teachers and the estimate of the statewide average calculated by KDE/ASME, based on the summer scoring -- are shown in Table 12-29. The first set of data is the statewide averages, according to the original scores reported by the teachers and then those same data after they have been recast by using the results from the summer rescoring. Results are shown for all three years of the program.

TABLE 12-29
RECAST WRITING PORTFOLIO RESULTS

Year	Grade 4		Grade 8		Grade 12	
	Original	Recast	Original	Recast	Original	Recast
1992	16.1	12.5	20.5	14.1	28.3	25.7
1993	32.7	19.6	29.7	18.2	42.9	27.6
1994	37.5	28.5	31.2	17.1	39.3	22.0

One observation is that recasting lowers scores in every case--that is, teachers have scored portfolios too high and continue to do so. Some interesting points are that the gap between original and recast scores narrowed this year in grade 4, but increased in grades 8 and 12, both in the original teacher judgements and in the results from the summer rescoring. The most striking finding is that recast writing portfolio scores are lower in 1994 than they were in 1992 for the 12th graders. Only in grade 4 did performance improve and were teachers getting closer to the intended standards with their scoring over the three years. However, a more detailed examination of the results has produced some startlingly positive results.

First , the correlations between original scores and rescores is improving. The correlations for 1993-94 at the student level were .70 at grade 4, .67 at grade 8, and .69 at grade 12. When school averages are computed, the correlations (for schools with 10 or more portfolio rescores only) are .70 for grade 4, .65 for grade 8, and .56 for grade 12. These correlations are not particularly high or encouraging, but they are higher than the lower-bound estimate of .50 given in the 1991-92 Technical Report. As one source of evidence of improvement, it is worth noting that the Scoring Analysis schools involved in the audit in 1992-93 had much more accurate scores in 1993-94 as a group than the other schools, as a group. Data in table 12-30 tell the story (these data are not recast, and therefore show means that are somewhat different from the ones provided above).

TABLE 12-30
1993-94 PORTFOLIO SCORE ACCURACY STATISTICS FOR SCHOOLS AUDITED IN 1992-93

Statistic	Grade 4		Grade 8		Grade 12	
	Audit	Non-Audit	Audit	Non-Audit	Audit	Non-Audit
Number of Schools	22	340	5	172	14	101
Number of Portfolios Rescored	401	6692	4954	138	3209	446
Percentage of Agreements with Summer Rescores	65	62	63	50	54	48
Correlation with Summer Rescores	.70	.70	.67	.67	.73	.69
Mean of Original Rescores	49.6	48.9	43.3	51.9	56.1	58.8
Mean of Summer Rescores	42.8	34.9	31.2	28.9	36.7	32.9
Difference between Original Scores and Summer Rescores	6.8	14.0	12.1	23.1	19.4	25.9

On every statistic, the schools with scoring differences larger than the "acceptable scoring range" in the 1992-93 writing portfolio audit scored better than the average school statewide. Most dramatically, these schools, which last year had scoring errors that were twice the statewide average, had half the statewide error this year at grades 4 and 8. This is a dramatic improvement. In one year, the least accurately scoring schools became among the most accurately scoring schools.

SECTION IV
GENERALIZABILITY

Section 4 examines issues related to test score validity in a broad sense. Chapter 13 presents reliability estimates for several components of the test. Chapter 14 describes existing validity-related evidence for KIRIS. Chapter 15 explores consequential validity for the assessment. Chapter 16 examines differences in performance at the student level by gender and ethnicity.

CHAPTER 13
RELIABILITY OF KIRIS ACCOUNTABILITY INDICES

STUDENT-LEVEL RELIABILITY

Although accountability decisions take place only at the school level, results are reported to parents at the student level. It is useful, therefore, to examine reliabilities at that level. Table 13-1 lists student-level coefficient alphas for open-response items for the first accountability cycle, computed by form for common and matrix items, and averaged. These figures are calculated using results from students who were eligible to complete open-response tests and who were present on the day of a testing, with absence being determined by having blank responses to all common and matrix items; blank responses by students who had attempted at least one question were counted as zero in the computation of coefficient alpha. The reason for not including the responses of absent students is that score reliability would be slightly over-estimated.

With the single exception of Grade 4 Science, open-response reliabilities increased from 1991-92 to 1992-93, a logical result of increasing the number of common items from three to five and the number of matrix items from one to two. The 1994-95 assessment includes an additional common item in grades 8 and 11.

Although the Department of Education advises against making student-level decisions based on individual test scores alone, open-response test reliabilities compare favorably with reliabilities from other tests used to make student-level decisions. Composite score reliability is comparable to reliabilities noted for the ACT Composite, and individual subject area reliabilities are similar to ACT and CTBS subject area reliabilities, as well as to reliabilities for a national certification test in accountancy.

As a caveat, it should be noted that using coefficient alpha probably underestimates score reliability insofar as item raw scores are the basis for the computation, whereas the fundamental scaling method employs a logistic model. The use of item response theory takes into account differences in item difficulty not reflected in the computational use of raw scores. A five item test having a single, relatively difficult item is likely to yield a diminished coefficient alpha in comparison with results from a comparable test with essentially uniform item difficulty.

Student-level reliability estimates for portfolios and performance events are considered in this Technical Manual and in previous technical reports in terms of scorer agreement rather than score reliability per se. Given that each student produces only one writing portfolio and, with limited exceptions, takes only one performance event, score reliability cannot be examined in the same manner as open-response items.

TABLE 13-1
OPEN-RESPONSE TEST RELIABILITIES BY SUBJECT AND GRADE

Grade	Subject	1991-92 Open-Response (three items, 12 possible score points)	1992-93 Open-Response (five items, 20 possible score points)	1993-94 Open-Response (five items, 20 possible score points)
4	Reading	.58	.77	.70
	Mathematics	.63	.72	.75
	Science	.63	.65	.73
	Social Studies	.64	.69	.76
	Composite	.88	.92	.93
8	Reading	.72	.83	.80
	Mathematics	.61	.79	.71
	Science	.62	.75	.66
	Social Studies	.74	.81	.83
	Composite	.89	.94	.92
12	Reading	.79	.85	.75
	Mathematics	.65	.81	.76
	Science	.72	.81	.76
	Social Studies	.77	.85	.85
	Composite	.92	.95	.94

SCHOOL-LEVEL RELIABILITY

The KIRIS accountability system requires the computation of an accountability index for each school and district in the state. The accountability index is the simple average of six indices. Five of the indices indicate performance in cognitive areas (reading, writing, mathematics, science and social studies) derived from results of Kentucky's three forms of testing: on-demand testing, performance events, and portfolios. The sixth index is a combination of noncognitive factors used to measure the effectiveness of a school. Detailed information about the calculation of these indices is provided in Chapter 10.

Whereas the 1992-93 Technical Report examined school-level reliability using an analysis of variance study in which tasks and students were considered fixed, a broader approach is to consider test items and students as either random or fixed. When both students and items are treated as random effects, the question being asked is, to what extent do the results obtained in a given year on the basis of the particular test items used that year and the particular students within each school in an accountability grade generalize to a theoretical pool of other tasks that might have been selected and to other students who might have attended an accountability grade? Preliminary results from one grade are given in table 13-2; further studies will follow in subsequent technical reports. In this analysis, Accountability Index data from 1992-93 and 1993-94 were examined treating students and tasks as fixed or random. Using a scaling procedure designed to place open-response items in an accountability index metric and a computational routine taking students into account, the following generalizability coefficients were obtained on samples of size 48 from schools having at least 48 students.

These prelimnary results indicate an adequate level of reliability in each year for grade twelve schools as small as 48 students. Dependability coefficients, where it is appropriate to consider such figures, are only slightly lower (.01 for 1993-94 data, .02 for 1992-93 data) than corresponding generalizability coefficients, still indicating an adequate level of reliability. Calculations employed in obtaining these preliminary figures assume a writing portfolio reliability of .65. Future generalizability analyses will consider sampling issues, and will provide projected reliabilities for larger and smaller sample sizes. Additional research of interest to school-level reliability includes a resampling study conducted by Kingston and Dings which examines the consistency of school accountability decisions (available from KDE).

TABLE 13-2
PRELIMINARY RESULTS: GENERALIZABILITY COEFFICIENTS TREATING STUDENTS AND TASKS AS FIXED OR RANDOM

Year	Grade	Students and Tasks Fixed	Tasks Fixed Students Random	Students Fixed Tasks Random	Students and Tasks Random
1992-93	12	.95	.91	.94	.89
1993-94	12	.92	.91	.91	.88

CHAPTER 14

VALIDITY-RELATED EVIDENCE

In a statement endorsed by national associations of psychologists, researchers and practitioners of education and educational measurement, the Standards for Educational and Psychological Testing[1] describes validity as "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores." The Standards also describes three broad, traditional categories of evidence necessary to support such inferences: content-related, criterion-related and construct-related evidence of validity. Evidence of these kinds for the use of KIRIS in evaluating student performance at the school level is presented below.

In the ten years since the Standards, traditional notions of validity have been supplemented with specific criteria for performance assessments (e.g., Frederiksen & Collins[2] Linn, Baker & Dunbar[3]) as well as the idea that the consequences of using a given test are an important aspect of validity. Validity addresses the issue of whether a test measures what it is intended to measure, or even more broadly, whether it has the effect it is intended to have. In a high-stakes accountability system such as KIRIS, it is that larger definition that is of paramount importance. This type of validity, termed "consequential validity," was advocated by Samuel Messick[4] of Educational Testing Service, who is greatly responsible for the wide use of the term.

ISSUES REGARDING CONSEQUENTIAL VALIDITY

The importance of the consequential validity of the test becomes obvious when one thinks of the alternatives. KIRIS exists because the Kentucky legislature believed that KIRIS would increase the likelihood of KERA's success by supporting (indeed, driving) changes in the classrooms of Kentucky. Thus, all decisions related to KIRIS ultimately must be considered in light of the question, "Will this change lead to better instruction and more real achievement?" If not, the change cannot be justified.

Traditional thinking about reliability and validity could lead to an assessment design contrary to the true purpose of KIRIS: supporting real change in the classroom. If the tests were highly reliable and valid in the traditional sense, but higher test scores reflected only superficial changes motivated solely by the goal of raising those test scores, KIRIS would not be fulfilling its obligation. Of course, if the tests could not demonstrate sufficiently high levels of traditional reliability and validity, people would probably not view them as fair, and therefore it is unlikely that the tests would have the desired impact on classroom practice. Therefore, high degrees of reliability and validity in the traditional sense are probably **necessary** conditions for high consequential validity, but they certainly are not **sufficient** conditions. That is, high consequential validity is probably not possible without high degrees of traditional reliability and validity, but high degrees of traditional reliability and validity cannot guarantee high consequential validity.

---

[1] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington: APA.

[2] Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.

[3] Linn, R. L., Baker, E. L. & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.

[4] Linn, Robert L., editor, *Educational Measurement, Third Edition*, MacMillan Publishing Co., 1989.

Eventually, the success of KIRIS will be judged on the extent to which classroom practices have improved throughout Kentucky. If teachers merely make modest changes in their teaching in order to raise test scores, and increases in KIRIS scores are not reflected in other measures of learning, KIRIS will not be judged successful no matter what reliability or validity coefficients it might be able to demonstrate. However, that judgment is far in the future; thus far, KIRIS has only made a small portion of its eventual impact.

THE ROLE OF RELIABILITY IN SELECTING KIRIS ASSESSMENTS

The designers of KIRIS attempted to balance the need for reliability with consequential validity. Deciding whether to include a given type of assessment was guided by the following principle: The overall Accountability Index must have sufficient reliability, but once that criterion has been met, the decision of whether to include an assessment, and if so, how much to weight it, was based on the assessment's presumed impact on the change in teaching.

This is one reason why so much importance is placed on improving the reliability of the portfolios in the assessment system. In traditional forms of measurement, classroom teachers are passive observers of the testing program. With portfolios, they are active participants. Teachers must understand fully both the content standards and the performance standards (what is to be learned and how well it must be learned) if their students are to develop quality portfolios and the teachers are to evaluate them. If portfolio reliability were high enough, it might be desirable to make portfolios the sole (or at least a substantial) factor on which the assessment system is based. They do not yet have reliability at the desired level, however, so they play a part, but only a part, in the assessment package.

TRADITIONAL VALIDITY-RELATED EVIDENCE, WITH NOVEL CRITERIA

Although consequential validity concerns may ultimately prove more important than issues of technical quality, it remains very important to examine KIRIS score validity from a traditional psychometric perspective. Thus, content validity information is reviewed below, and construct-related validity evidence is discussed on the basis of correlations of KIRIS tests with each other and with scores from other testing programs. Intermingled with traditional notions of validity in this analysis are more recently proposed criteria for evaluating performance assessments: systematic validity, directness and transparency[5]; and, fairness, transfer, generalizability, cognitive complexity, content quality, content coverage, meaningfulness, cost, efficiency, and consequences[6].

CONTENT-RELATED VALIDITY EVIDENCE

Section II of this manual describes how the components of the KIRIS assessment were derived from Kentucky learner outcomes and academic expectations, using advisory committees of Kentucky teachers to make those outcomes and expectations operational through test items, and to make choices about what the tests would contain. Several tables summarize the learner outcomes and academic expectations measured by the items in the assessment, and there is little to add to the extensive treatment of this material. In short, there is substantive validity-related evidence in the process by which KIRIS assessments were constructed.

---

[5] Frederiksen & Collins.

[6] Linn, Baker & Dunbar.

While test development information serves as the primary source of content-related validity evidence, examining KIRIS tests in terms of the novel content-relevant criteria noted above provides a potential source of additional evidence. Cognitive complexity, content quality, and content coverage can serve as criteria by which to evaluate performance assessments. Because there exist no established standards for these criteria (as noted by Linn, Baker and Dunbar), any detailed consideration of them probably requires discourse substantiated by expert judgment in the form of task analysis. As noted in Section II, content coverage by open-response items in KIRIS was found to be acceptable as judged against the coverage afforded by multiple-choice items for the particular number of each type of question administered in 1993-94. Although content quality *per se* has not been examined, the presence of teacher and other content area specialist judgment in writing and selecting items for the assessments provides some indication of validity in this regard. Arguably, most performance events embody essential characteristics of cognitively complex tasks, requiring the student to demonstrate problem-solving.

CONSTRUCT-RELATED VALIDITY EVIDENCE

RELATIONSHIPS AMONG KIRIS MEASURES. For a testing program assessing student performance in several content areas using a variety of testing methods, the empirical relationships among performances in each content area and on the various methods of assessment can provide evidence supporting the inferences made from test scores. The most convenient summary of much of the potential validity-related evidence provided by these relationships is a correlation matrix, laid out in a manner analogous to the multitrait, multimethod matrix first described by Campbell and Fiske[7]. For KIRIS, the constructs of interest are not multiple traits but rather performances in several content areas, assessed using Performance Events, Portfolios, On-Demand Writing Prompts, and Open-Response and Multiple-Choice tests. The lower diagonals of tables 14-1, 14-2, and 14-3 contain correlations at the student level among multiple-choice and open-response common item raw scores (0 to 20 points), and writing prompt and portfolio scores on a four-point score scale (1 through 4 points) corresponding to the Novice, Apprentice, Proficient, and Distinguished levels of performance; the upper diagonals contain correlations which have been disattenuated based on the reliability estimates found on the diagonal. Only students having a complete set of scores were included in this analysis. (Performance event scores were tracked at the school level rather than the individual level, and therefore were not available for this analysis.)

From a traditional perspective, correlation-based validity evidence regarding content area and method of assessment resides in positive, relatively high correlations among scores in the same content area measured using different methods. For example, finding high correlations among portfolio, open-response and multiple-choice measures of mathematics performance would be considered favorable, especially if the correlations among these measures were higher than their correlations with measures of reading, science and social studies. In the case of KIRIS, however, the performance-based methods of open-response testing and portfolio assessment are designed to require performances that are related to but not identical to the performances required by most multiple choice tests. From this perspective, it is reasonable to expect that some indication of method effect will be found, as evidenced by moderately high correlations within each method of assessment across different content areas.

---

[7] Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

TABLE 14-1
GRADE 4 STUDENT-LEVEL CORRELATIONS AMONG COMMON ITEM SCORES (N=47,851)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio	
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math
Multiple-Choice	Reading	***.83***	.77	.86	.88	.77	.72	.75	.67	.43	.72	.63
	Math	.61	***.76***	.86	.86	.65	.85	.70	.62	.42	.70	.67
	Science	.66	.63	***.71***	.92	.69	.78	.78	.68	.40	.70	.63
	Soc. St.	.69	.62	.67	***.74***	.70	.75	.75	.71	.42	.71	.65
Open-Response	Reading	.62	.49	.51	.53	***.78***	.77	.80	.79	.52	.79	.70
	Math	.57	.64	.57	.56	.59	**.75**	.84	.78	.49	.79	.76
	Science	.58	.52	.56	.55	.60	.62	***.73***	.82	.46	.75	.67
	Soc. St.	.53	.48	.50	.53	.61	.59	.61	***.76***	.64	.82	.70
On-Demand	Writing	.28	.26	.24	.26	.33	.30	.28	.30	***.51***	.64	.58
Portfolio	Writing	.46	.43	.41	.43	.49	.48	.45	.46	.32	***.49***	1.20
	Math	.40	.41	.37	.39	.43	.46	.40	.41	.29	.59	***.49***

TABLE 14-2
GRADE 8 STUDENT-LEVEL CORRELATIONS AMONG COMMON ITEM SCORES (N=48,209)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio	
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math
Multiple-Choice	Reading	***.82***	.70	.78	.84	.82	.69	.73	.68	.50	.81	.73
	Math	.56	***.78***	.88	.78	.61	.86	.72	.59	.45	.76	.82
	Science	.54	.59	***.58***	.89	.68	.76	.81	.65	.45	.78	.76
	Soc. St.	.67	.61	.60	***.78***	.71	.73	.79	.71	.48	.81	.74
Open-Response	Reading	.66	.48	.46	.56	***.79***	.75	.78	.80	.56	.86	.67
	Math	.53	.64	.49	.54	.56	***.71***	.79	.72	.50	.83	.74
	Science	.53	.50	.50	.56	.56	.54	***.65***	.79	.51	.80	.67
	Soc. St.	.56	.47	.45	.57	.64	.55	.58	***.82***	.54	.82	.67
On-Demand	Writing	.37	.33	.28	.35	.41	.35	.34	.40	***.68***	.76	.70
Portfolio	Writing	.49	.45	.40	.48	.51	.47	.43	.49	.42	***.45***	1.25
	Math	.40	.44	.35	.40	.42	.45	.38	.43	.35	.51	***.37***

TABLE 14-3
GRADE 12 STUDENT-LEVEL CORRELATIONS AMONG COMMON ITEM SCORES (N=33,612)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio	
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math
Multiple-Choice	Reading	***.76***	.68	.78	.82	.77	.66	.74	.68	.52	.79	.54
	Math	.51	***.73***	.88	.77	.63	.88	.76	.61	.50	.76	.73
	Science	.59	.65	***.75***	.89	.64	.80	.83	.67	.49	.75	.64
	Soc. St.	.61	.56	.66	***.73***	.69	.70	.77	.73	.51	.78	.58
Open-Response	Reading	.58	.46	.48	.51	***.74***	.75	.81	.80	.61	.86	.64
	Math	.50	.65	.60	.52	.56	***.75***	.84	.73	.55	.82	.82
	Science	.56	.56	.62	.57	.60	.63	***.75***	.81	.58	.80	.67
	Soc. St.	.54	.48	.53	.57	.63	.58	.64	***.84***	.56	.79	.62
On-Demand	Writing	.38	.36	.36	.37	.44	.40	.42	.43	***.71***	.74	.55
Portfolio	Writing	.48	.45	.45	.46	.51	.49	.48	.50	.43	***.48***	.88
	Math	.33	.44	.39	.35	.39	.50	.41	.40	.33	.43	***.50***

Every correlation between a multiple-choice and open-response measure is positive and indicates at least a moderate degree of relationship. The .45 correlation of Grade 8 Open-Response Social Studies with Multiple-Choice Science is the lowest. Of the correlations between Open-Response and Multiple-Choice test scores, nine of the twelve Open-Response tests (four content areas at three grade levels) have their highest correlation with a Multiple Choice test having the same content area. While the differences between these nine correlations and the next-highest of the Open-Response-Multiple-Choice correlations involving the same Open-Response test are small to moderate in magnitude (ranging from .01 to .10), the pattern found among the correlations is fairly consistent, and differences as small as .01 are statistically significant at the .01 level for the sample sizes employed. The exceptions to this trend are that Grade 4 Open-Response Science correlates somewhat more highly with Multiple-Choice Reading than with Multiple-Choice Science, Grade 4 Open-Response Social Studies correlates about as highly with Multiple-Choice Reading as with Multiple Choice Social Studies, and Grade 8 Open-Response Science correlates with Multiple-Choice Reading, Mathematics and Social Studies more highly than with Multiple-Choice Science. It is also worth noting that grade 12 Open-Response Social Studies correlates about as highly with Multiple-Choice Social Studies as Multiple-Choice Social Studies correlates with Open-Response Science. Although it is important to avoid over-interpretation of small differences in correlations, these findings support score validity to the extent that the strongest relationships across measures tend to occur within a content area.

The Multiple-Choice data suggest the presence of a method effect more strongly than do the Open-Response data. Except for Mathematics at grades 4 and 8, every Multiple-Choice test correlates at least as highly with another Multiple-Choice test (in a different content area) as it does with the Open-Response test in its same content area. Likewise, Open-Response tests in Social Studies and Science at each grade level correlate more highly with another Open-Response test (in a different content area) than with the Multiple-Choice test in their same content area. By contrast, Open-Response Reading and Mathematics show their highest correlations in most cases with Multiple-Choice Reading and Mathematics, respectively;

the sole exception is Grade 12 Reading. This suggests that Reading and Mathematics are operationalized more consistently across the two testing methods than are Science and Social Studies. One potential explanation for this finding is that the relatively deep but somewhat narrow content coverage afforded by five-question Open-Response testing has a greater impact on Science and Social Studies than on Reading and Mathematics. Another possibility, although less clearly supported by the data, is that the writing ability required to answer essay-type questions plays a moderating role in students' performances across content areas requiring extended written answers. In any case, the lowest correlation found between two different content area tests employing the same method is .51 (Grade 12 Multiple-Choice Reading with Mathematics).

Writing and Mathematics Portfolio score correlations with Multiple-Choice and Open-Response scores produce consistent, interpretable patterns of results. For every pair of Open-Response and Multiple-Choice tests in a given content area, Portfolios correlate with Open-Response tests somewhat more highly than with Multiple-Choice tests. These correlations, which range in magnitude from .33 for Grade 12 Mathematics Portfolio with Multiple-Choice Reading to .51 for Writing with Open-Response Reading at grades 8 and 12, suggest that the performance tapped by Portfolio assessment is more similar to that measured by Open-Response than by Multiple-Choice, but not identical to that of Open-Response. This observed relationship provides validity evidence by being logically consistent with the idea that performance-based Portfolio assessment should be more similar to constructed-response tests (such as Open-Response) than to Multiple-Choice.

An apparent content area relationship exists for Portfolios. At all three grade levels, Mathematics Portfolios correlate somewhat more highly with Open-Response Mathematics than with other Open-Response tests, and somewhat more highly with Multiple-Choice Mathematics than with other Multiple-Choice tests. This validity evidence supports the idea that mathematics performances are being measured in each case. To the extent that performance of a writing task should be more highly related to performance in reading than to performances in the other content areas, correlations involving Writing Portfolios lend support to Writing Portfolio score validity. In nearly every case, Writing Portfolios correlate somewhat more highly with Open-Response Reading than with other Open-Response tests, and somewhat more highly with Multiple-Choice Reading than with other Multiple-Choice tests. Correlations between Writing Portfolio scores and On-Demand Writing scores provide somewhat favorable validity-related evidence, with higher correlations in grades 8 and 12 ($r = .42$ and $r = .43$, respectively) than in grade 4 ($r = .33$). A lack of variability in the On-Demand Writing scores, especially at grade 4, is responsible to some extent for the low magnitude of the correlation with Writing Portfolio scores.

Evidence of a method effect for Portfolios varies somewhat across grades. At grade 4, Writing and Mathematics Portfolios correlate more highly with each other ($r = .59$) than with any other measure. At grade 8, Writing and Mathematics Portfolios correlate as highly with each other ($r = .51$) as Writing Portfolio scores correlate with Open-Response Reading. At grade 12, however, Writing Portfolio scores correlate somewhat more highly with Open-Response and Multiple-Choice tests in every content area than with Mathematics Portfolios ($r = .44$). Grade 12 Mathematics Portfolio scores correlate at least as highly with Open-Response and Multiple-Choice tests in Mathematics ($r = .49$ and $r = .44$, respectively) as with Writing Portfolio scores. This suggests a method effect explained to some degree by parental influence and the role of student effort in creating and revising portfolio entries. In grade 4, the same teacher is likely to coach a student's portfolio endeavors in mathematics and writing. The influence of the same teacher on each student's performance is likely not a factor in grades eight and twelve, where mathematics and writing are typically addressed by different teachers. It is also likely that parental influence is greater in grade 4 than in grades 8 and 12. In addition, the role of writing in creating a Mathematics Portfolio might contribute to the relationship between Writing and Mathematics Portfolios. And while it is important to consider that correlations are influenced by the distributions of scores in the variables upon which they are based, the relationship between mathematics and reading portfolios at grade 4 does not appear to be artificially inflated by the extent to which portfolio score distributions are dominated by the lowest level of achievement.

The student-level correlations described above are worth considering in several other ways as well: computed with scores of students who appear not to have made a good faith effort removed; and, as disattenuated correlations, corrected for estimated unreliability of the underlying measures. These correlations are examined below, along with school-level correlations for which performance events data is available.

DISATTENUATED CORRELATIONS. Disattenuated correlations are presented for grades 4, 8, and 12 in the upper right hand corner of tables 14-1, 14-2 and 14-3, with reliability estimates appearing on the diagonal. Coefficient alpha was used to estimate reliability for Multiple-Choice and Open-Response assessments. Portfolio reliability estimates are squared correlation coefficients between two sets of scores on a subset of portfolios, as described in Chapter 12[8], while on-demand writing reliability estimates reflect squared correlation coefficients between scores on pieces evaluated twice as part of the quality control procedures followed for these assessments.

A disattenuated correlation is computed by dividing a correlation coefficient by the square root of the product of the reliability estimates for the measures upon which the correlation is based. The logic behind this computation is that if two tests measure the same construct, but do so imperfectly, their disattenuated correlation should approach unity. However, a disattenuated correlation approaching or even exceeding 1.00 should only be taken as an indication of the possibility that the same construct or two strongly related constructs are being measured. Correlational evidence is not causal evidence of a relationship, and is specific to the sample used in computing the correlations.

Only one pair of assessments yields a disattenuated correlation at or above 1.00: mathematics portfolio with writing portfolio for grade 4 and grade 8. The disattenuated correlation between these two measures is relatively high for grade 12 ($r=.88$). Finding such high disattenuated correlations between two evidently distinct measures is probably the result of using a very conservative reliability estimate for portfolio scores (i.e., the squared correlation between two sets of scores given to the same portfolios by different raters). In addition, the role of student effort, writing, and, in grade 4, teacher influence, may provide some explanation of this relationship, as described above.

Among multiple-choice and open-response tests, for which coefficient alpha served as the reliability estimate, the only remarkable results occur for multiple-choice Science. Multiple-choice Science scores correlate .92, .89, and .89 with multiple-choice social studies at grades 4, 8, and 12, respectively, and correlate only slightly lower with multiple-choice mathematics. In addition to the method effect suggested by these findings, the disattenuated correlations evidence essentially the same patterns described above for the correlations not adjusted for unreliability.

STUDENT-LEVEL CORRELATIONS WHEN LACK OF GOOD FAITH EFFORT SCORES ARE REMOVED. One general difficulty in interpreting test results is that some students do not make a reasonable attempt to perform well. Such students are likely to score low on every component of the KIRIS assessments, and their inclusion in this sample is likely to increase correlations among the various components spuriously. As an additional analysis, correlations were computed after scores were removed for students having a score of three or less on a twenty-item multiple-choice test or on an open response test with a maximum score of twenty points. (The score of three or less was selected somewhat arbitrarily as being more than one point lower that chance level on a four-option multiple-choice test, as well as being more than one point below the score obtained by producing anything other than a off-topic response or a blank response to each of the five items of the open-response test.) These correlations are presented in tables 14-4, 14-5, and 14-6.

---

[8] Because no data were available from any Grade 4 Mathematics Portfolio quality control efforts, no direct reliability estimate was available for this measure. Instead, this reliability was estimated using the squared correlation for grade 4 Writing Portfolios.

Not surprisingly, correlations computed after removing students whose scores suggested the lack of a good faith effort were lower than the original correlations in every instance. Correlations involving portfolio scores generally decreased by a smaller amount than the other correlations; this result is not surprising, given that no standard for a lack of good faith effort could be applied to portfolio scores to remove a student's data from this analysis. In any case, the pattern of correlations resulting from this analysis did not suggest any new interpretations with regard to testing method or content area.

SCHOOL-LEVEL CORRELATIONS. School-level correlations, presented in tables 14-7, 14-8, and 14-9, suggest a method effect for Multiple-Choice and Open-Response tests. For almost every grade and content area combination, correlations among Open-Response tests and among Multiple-Choice tests are greater than the correlation between any pair of Open-Response and Multiple-Choice tests assessing the same content area; the exceptions are found in grade 8, where Reading (r=.69) and Science (r=.92) correlate more highly across assessment methods than Open-Response Mathematics correlates with Open-Response Reading (r=.63) and/or Social Studies (r=.61). In fact, grade 8 provides the only instances of Open-Response to Open-Response correlations being lower than any of the Open-Response to Multiple-Choice correlations. There are no examples of Multiple-choice to Multiple-choice correlations being lower than an Open-Response to Multiple-Choice correlation.

TABLE 14-4
GRADE 4 STUDENT-LEVEL CORRELATIONS, LACK OF GOOD FAITH EFFORT SCORES REMOVED (N=38,053)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio	
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math
Multiple-Choice	Reading	*1.00*	.52	.58	.62	.50	.44	.45	.40	.24	.40	.34
	Math	.52	*1.00*	.54	.53	.37	.55	.40	.35	.21	.36	.34
	Science	.58	.54	*1.00*	.60	.38	.46	.45	.38	.20	.33	.30
	Soc. St.	.62	.53	.60	*1.00*	.41	.45	.44	.42	.22	.36	.33
Open-Response	Reading	.50	.37	.38	.41	*1.00*	.48	.47	.50	.30	.42	.37
	Math	.44	.55	.46	.45	.48	*1.00*	.49	.48	.26	.40	.39
	Science	.45	.40	.45	.44	.47	.49	*1.00*	.48	.24	.36	.33
	Soc. St.	.40	.35	.38	.42	.50	.48	.48	*1.00*	*.27*	.38	.34
On-Demand	Writing	.24	.21	.20	.22	.30	.26	.24	.27	*1.00*	.30	.26
Portfolio	Writing	.40	.36	.33	.36	.42	.40	.36	.38	.30	*1.00*	.56
	Math	.34	.34	.30	.33	.37	.39	.33	.34	.26	.56	*1.00*

TABLE 14-5
GRADE 8 STUDENT-LEVEL CORRELATIONS, LACK OF GOOD FAITH EFFORT SCORES REMOVED (N=30,573)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio	
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math
Multiple-Choice	Reading	***1.00***	.47	.44	.58	.50	.35	.36	.41	.29	.41	.34
	Math	.47	***1.00***	.53	.54	.36	.54	.38	.35	.25	.37	.38
	Science	.44	.53	***1.00***	.53	.32	.37	.38	.32	.20	.31	.28
	Soc. St.	.58	.54	.53	***1.00***	.42	.38	.42	.43	.27	.39	.33
Open-Response	Reading	.50	.36	.32	.42	***1.00***	.39	.39	.50	.34	.43	.37
	Math	.35	.54	.37	.38	.39	***1.00***	.37	.39	.24	.36	.36
	Science	.36	.38	.38	.42	.39	.37	***1.00***	.42	.24	.32	.28
	Soc. St.	.41	.35	.32	.43	.50	.39	.42	***1.00***	.34	.41	.37
On-Demand	Writing	.29	.25	.20	.27	.34	.24	.24	.34	***1.00***	.37	.30
Portfolio	Writing	.41	.37	.31	.39	.43	.36	.32	.41	.37	***1.00***	.47
	Math	.34	.38	.28	.33	.37	.36	.28	.37	.30	.47	***1.00***

TABLE 14-6
GRADE 12 STUDENT-LEVEL CORRELATIONS, LACK OF GOOD FAITH EFFORT SCORES REMOVED (N=22,730)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio	
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math
Multiple-Choice	Reading	***1.00***	.44	.50	.53	.46	.36	.43	.42	.31	.41	.28
	Math	.44	***1.00***	.62	.52	.36	.60	.48	.40	.29	.39	.40
	Science	.50	.62	***1.00***	.60	.36	.52	.53	.42	.29	.38	.34
	Soc. St.	.53	.52	.60	***1.00***	.40	.42	.48	.49	.31	.40	.31
Open-Response	Reading	.46	.36	.36	.40	***1.00***	.42	.47	.51	.38	.44	.34
	Math	.36	.60	.52	.42	.42	***1.00***	.50	.46	.31	.40	.46
	Science	.43	.48	.53	.48	.47	.50	***1.00***	.50	.34	.39	.36
	Soc. St.	.42	.40	.42	.49	.51	.46	.50	***1.00***	.36	.42	.34
On-Demand	Writing	.31	.29	.29	.31	.38	.31	.34	.36	***1.00***	.38	.28
Portfolio	Writing	.41	.39	.38	.40	.44	.40	.39	.42	.38	***1.00***	.40
	Math	.28	.40	.34	.31	.34	.46	.36	.34	.28	.40	***1.00***

TABLE 14-7
GRADE 4 SCHOOL-LEVEL CORRELATIONS (N=801)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio		Performance Events		
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math	Math	Science	Soc. St.
Multiple-Choice	Reading	***1.00***	.78	.78	.83	.61	.58	.60	.52	.56	.40	.30	.41	.35	.42
	Math	.78	***1.00***	.80	.81	.54	.65	.61	.54	.53	.39	.32	.38	.34	.38
	Science	.78	.80	***1.00***	.85	.56	.61	.66	.56	.53	.38	.31	.34	.33	.34
	Soc. St.	.83	.81	.85	***1.00***	.55	.59	.62	.58	.54	.38	.31	.37	.36	.38
Open-Response	Reading	.61	.54	.56	.55	***1.00***	.76	.78	.78	.56	.46	.34	.26	.24	.28
	Math	.58	.65	.61	.59	.76	***1.00***	.84	.79	.53	.47	.42	.35	.30	.33
	Science	.60	.61	.66	.62	.78	.84	***1.00***	.82	.53	.47	.36	.28	.27	.29
	Soc. St.	.52	.54	.56	.58	.78	.79	.82	***1.00***	.53	.44	.35	.27	.26	.27
On-Demand	Writing	.56	.53	.53	.54	.56	.53	.53	.53	***1.00***	.44	.34	.30	.31	.35
Portfolio	Writing	.40	.39	.38	.38	.46	.47	.47	.44	.44	***1.00***	.61	.22	.22	.23
	Math	.30	.32	.31	.31	.34	.42	.36	.35	.34	.61	***1.00***	.21	.22	.21
Performance Events	Math	.41	.38	.34	.37	.26	.35	.28	.27	.30	.22	.21	***1.00***	.47	.46
	Science	.35	.34	.33	.36	.24	.30	.27	.26	.31	.22	.22	.47	***1.00***	.44
	Soc. St.	.42	.38	.34	.38	.28	.33	.29	.27	.35	.23	.21	.46	.44	***1.00***

TABLE 14-8
GRADE 8 SCHOOL-LEVEL CORRELATIONS (N=368)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio		Performance Events		
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math	Math	Science	Soc. St.
Multiple-Choice	Reading	***1.00***	***.78***	.76	.79	.63	.45	.58	.55	.65	.39	.29	.55	.35	.42
	Math	.78	***1.00***	.79	.77	.51	.56	.57	.48	.61	.40	.36	.59	.39	.34
	Science	.76	.79	***1.00***	.82	.41	.38	.59	.45	.56	.33	.29	.58	.40	.40
	Soc. St.	.79	.77	.82	***1.00***	.46	.36	.55	.54	.56	.34	.26	.54	.37	.40
Open-Response	Reading	.63	.51	.41	.46	***1.00***	.61	.69	.78	.67	.47	.28	.37	.29	.33
	Math	.45	.56	.38	.36	.61	***1.00***	.64	.58	.48	.40	.33	.31	.23	.27
	Science	.58	.57	.59	.55	.69	.64	***1.00***	.71	.62	.41	.28	.42	.37	.35
	Soc. St.	.55	.48	.45	.54	.78	.58	.71	***1.00***	.62	.44	.24	.36	.34	.37
On-Demand	Writing	.65	.61	.56	.56	.67	.48	.62	.62	***1.00***	.53	.36	.42	.36	.38
Portfolio	Writing	.39	.40	.33	.34	.47	.40	.41	.44	.53	***1.00***	.44	.24	.25	.24
	Math	.29	.36	.29	.26	.28	.33	.28	.24	.36	.44	***1.00***	.25	.23	.18
Performance Events	Math	.55	.59	.58	.54	.37	.31	.42	.36	.42	.24	.25	***1.00***	.35	.46
	Science	.35	.39	.40	.37	.29	.23	.37	.34	.36	.25	.23	.35	***1.00***	.39
	Soc. St.	.42	.34	.40	.40	.33	.27	.35	.37	.38	.24	.18	.46	.39	***1.00***

TABLE 14-9
GRADE 12 SCHOOL-LEVEL CORRELATIONS (N=238)

Test	Content	Multiple-Choice				Open-Response				On Demand	Portfolio		Performance Events		
		Reading	Math	Science	Soc. St.	Reading	Math	Science	Soc. St.	Writing	Writing	Math	Math	Science	Soc. St.
Multiple-Choice	Reading	*1.00*	.83	.86	.86	.56	.52	.61	.53	.64	.38	.34	.49	.42	.33
	Math	.83	*1.00*	.88	.87	.59	.67	.63	.59	.71	.44	.42	.52	.48	.34
	Science	.86	.88	*1.00*	.89	.58	.59	.66	.56	.72	.41	.36	.46	.49	.35
	Soc. St.	.86	.87	.89	*1.00*	.56	.60	.62	.58	.67	.42	.38	.41	.47	.34
Open-Response	Reading	.56	.59	.58	.56	*1.00*	.77	.83	.87	.77	.41	.28	.39	.41	.36
	Math	.52	.67	.59	.60	.77	*1.00*	.79	.81	.66	.36	.39	.37	.36	.26
	Science	.61	.63	.66	.62	.83	.79	*1.00*	.86	.71	.33	.34	.36	.43	.35
	Soc. St.	.53	.59	.56	.58	.87	.81	.86	*1.00*	.71	.37	.29	.39	.38	.30
On-Demand	Writing	.64	.71	.72	.67	.77	.66	.71	.71	*1.00*	.49	.35	.42	.49	.37
Portfolio	Writing	.38	.44	.41	.42	.41	.36	.33	.37	.49	*1.00*	.29	.37	.29	.25
	Math	.34	.42	.36	.38	.28	.39	.34	.29	.35	.29	*1.00*	.30	.30	.20
Performance Events	Math	.49	.52	.46	.41	.39	.37	.36	.39	.42	.37	.30	*1.00*	.37	.23
	Science	.42	.48	.49	.47	.41	.36	.43	.38	.49	.29	.30	.37	*1.00*	.48
	Soc. St.	.33	.34	.35	.34	.36	.26	.35	.30	.37	.25	.20	.23	.48	*1.00*

Correlations among performance events more strongly suggest a method effect in grade 4 than in the other grades. However, even these correlations (in the mid .40s for grade 4) are influenced by the scoring of inter-disciplinary performance events. Of the three mathematics performance events contributing to scores in grade 4, for example, one is unique to mathematics, one is counted toward social studies as well as mathematics, and another is counted toward science. A performance event is shared between social studies and science as well. Although these items are scaled somewhat differently across content areas, the same performance serves as the basis for the scoring. (While it may be argued that such a scoring procedure is appropriate to performances which are inter-disciplinary in nature, this practice is likely to inflate the correlations.) Beyond this suggestion of a method effect, the most remarkable aspect of correlations involving performance events is that they tend to be rather low. The magnitude of the correlations is certainly influenced by the sampling error associated with each student taking one of thirteen to sixteen performance events.

CONCURRENT VALIDITY

Another measure of the validity of a test is how well it correlates to other accepted measures of the same or similar constructs. Because no other "primarily performance-based" assessments exist for comparison with KIRIS, options are limited for demonstrating the concurrent validity of the test. Even if other such measures existed, it is unlikely that they would measure the same constructs as KIRIS. The best one can do is to compare the performance of students on KIRIS to other traditionally accepted tests of academic achievement. In this case, one hopes that the correlations will be high, since much of what is measured by most testing programs will be in common, but not too high. If the correlations are very high, it would imply that whatever is supposed to be new in KIRIS is either missing or highly correlated with traditional forms of assessment. If that were the case, KIRIS could scarcely be justified, since the traditional forms of assessment are usually so much less expensive and time-consuming to administer.

Another reason for comparing KIRIS to traditional forms of assessment is that those traditional measures are still in use. For example, a large number of Kentucky high school students will continue to take the American College Testing programs (ACTs) for some time to come for college admissions. If KIRIS proved to be uncorrelated with that measure, it would place students, parents, and teachers in the uncomfortable position of having to choose the test on which they would like to focus their attention. Fortunately, as will be shown in the ensuing discussion, there is significant correlation between KIRIS and ACT scores, and it is not overly presumptuous to assume that increased learning that leads to improvement on one is likely to lead to improvement on the other.

Yet another reason for linking current assessment measures to those of the past is that it gives a sense of continuity and relationship. The Comprehensive Test of Basic Skills (CTBS) was given widely in Kentucky before the inception of KIRIS, and it would be disquieting if there were little or no relationship between the two measures. Again, there is a high degree of relationship, and for those who question how *avant garde* KIRIS is, it will be reassuring to them to know that the correlations between the old and new are sufficiently high to show that while the KIRIS transitional tests are a step in a new direction, they clearly are not too far a leap from traditional testing.

CORRELATIONS BETWEEN KIRIS AND OTHER TESTS. Studies have been completed to date that examine the relationship between the KIRIS transitional tests and three other tests: the Comprehensive Test of Basic Skills (CTBS), the National Assessment of Educational Progress (NAEP), and the American College Testing Program (ACT). Discussion of each of those studies follows.

THE COMPREHENSIVE TEST OF BASIC SKILLS. Administration of the CTBS (CTBS/4) was mandated for all students statewide at grades K-3, 5, 7, and 10 for the 1988-89 and 1989-90 school years. Testing was optional at the other grade levels in those years, and at all grades for the 1990-91 and 1991-92 school years. While testing was not mandated, virtually all districts statewide administered the test anyway, because the Kentucky Department of Education scored the tests for free. Administration of the CTBS was so prevalent that even in 1991-92, the first year for KIRIS, at least 24,000 students were tested at every grade level, even the accountability grades, except for grade 12. This meant that there were large numbers of students taking both tests at each grade level, leading to a natural study of the relationship between the two tests. Another consequence of the Kentucky Department of Education scoring the tests for free was that the Department had the data for virtually every Kentucky student tested on the CTBS during the 1991-92 school year. This data file was provided to Advanced Systems to do the analyses for this study.

To determine the correlations, students were matched from the two files (KIRIS and CTBS) on the basis of school code, the first six letters of their last name, and the first three letters of their first name. The reason for using a limited number of letters in each name was to reduce the number of mismatches due to miscoding of names. Matching on the basis of names, especially first names, is troublesome because, in addition to the normal problems that can occur with miscoding, students often use formal names on some forms and nicknames on others. By limiting the number of letters in the match, this problem is somewhat reduced.

The variables obtained from the matched file were the CTBS total reading and math Normal Curve Equivalents and the total open-response scores on the corresponding KIRIS measures. For 1991-92, the KIRIS open-response scores were based on three items at the accountability grades, and four items at the remaining grades. The correlations between the two tests are shown in Table 14-10.

The correlations for reading ranged from .54 to .65, except for the two accountability grades, where the correlations were .47. For mathematics, the correlations ranged from .47 in grade two to .65 in grade three. For most grades, the correlations were above .6. Given the limited number of questions on the KIRIS tests, these data show a modest correlation between the two tests. The coefficient alphas for the grade 4 tests in reading and math were reported in the *1991-92 Technical Report* as .59 and .73 respectively, and .66 and .63 for grade 8. If the reliability of the CTBS tests is .85, then the disattenuated correlations between KIRIS and CTBS would be .66 and .77 for grade 4 reading and math, respectively, and .63 and .86 for grade 8. These correlations indicate that the KIRIS tests are measuring something substantially different from the CTBS for reading, but that there is considerably more overlap for mathematics. For the 1992-93 KIRIS tests, the number of common open-response questions was increased from three to five for the accountability grades, so the reliability of student level data increased significantly. It is presumed, therefore, that there was a significant increase in the correlation between the KIRIS tests and the CTBS. Widespread administration of the CTBS stopped in 1992-93, however, so data of the type used for the 1991-92 study are unavailable to examine this question.

TABLE 14-10
CORRELATIONS BETWEEN CTBS NCES AND KIRIS OPEN-RESPONSE QUESTIONS, 1991-92

Grade	Reading		Mathematics	
	Number Matched	Correlation	Number Matched	Correlation
2	5,829	.65	5,623	.47
3	9,137	.62	8,676	.65
4	31,768	.47	30,688	.61
5	5,040	.54	4,911	.64
6	3,938	.54	3,846	.61
7	8,739	.54	8,528	.63
8	12,293	.47	12,253	.63
9-11	11,920	.55	11,813	.57

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS. For years, the NAEP revealed data only to regional levels. (NAEP divided the United States into four regions for purposes of reporting.) Beginning with the 1990 assessment, however, NAEP began its Trial State Assessment (TSA). Through the TSA, a state could request to have an additional sampling of schools taken within the state and receive information on the performance of its own students. For 1990, this program was limited to grade 8 mathematics. For the 1992 assessment, this program was expanded to include reading and mathematics at grade 4.

Realizing the potential power of this information, Kentucky requested that NAEP conduct a study in the state that went beyond the grades and content areas assessed by the TSA, and that additional samples be drawn so that Kentucky would have data for grades 4, 8, and 12 in reading, mathematics, and writing. NAEP agreed to draw the samples and test this additional group of students.

The primary purpose for this additional study was to determine the relationship between KIRIS standards and those being set by NAEP, and to establish baselines of Kentucky performance so that NAEP could be used in the future as an external measure of Kentucky's progress. Once the data had been collected, however, it became evident that it also would serve as a measure of the commonality of KIRIS and NAEP. However, NAEP does not collect student names with its data. Therefore, no correlation could be computed between the two tests at the student level. NAEP did provide a school code with each student's record, however, so it was possible to compute school averages and correlate school mean scores on the two tests.

NAEP tested only a sample of students in each school, so that while the KIRIS data were based on all students in the accountability grades, NAEP's data for school means are based on a limited number (9-31 students for grade 4, 16-31 for grade 8, and 14-45 in grade 12, with the exception of one school that tested 87 students). The number of NAEP schools and students for which matching data from KIRIS were available is shown in Table 14-11, along with the correlations between the NAEP mean scaled score for each school and the school's mean score for the KIRIS open-response questions. Because NAEP tested such a limited number of students in each school, it is likely that the school-level correlations are significantly lower than if all students in each school had been tested.

TABLE 14-11
NUMBERS OF SCHOOLS AND STUDENTS OBTAINED FROM NAEP SAMPLING, AND CORRELATION BETWEEN NAEP AND KIRIS MATHEMATICS

Grade	Number of Students	Number of Schools	Correlation between NAEP and KIRIS
4	2,585	111	.74
8	2,702	102	.78
12	997	37	.79

As noted above, the size of these correlations is limited by the number of students tested by NAEP in each school. Also, just as was true for the CTBS study, these results are based on the 1991-92 KIRIS testing. The 1992-93 KIRIS testing included an increased number of open-response questions, which should lead to higher correlations than the ones shown above.

Interestingly, NAEP administers a mixture of open-response and multiple-choice questions. As a point of comparison, the scores of the KIRIS multiple-choice questions were correlated with the NAEP scaled scores. The correlations for grades 4, 8, and 12 were .72, .63, and .81, respectively, which means that the correlation between NAEP and the KIRIS multiple-choice questions was about the same as the correlation with KIRIS open-response questions for grades 4 and 12, and considerably less for grade 8. That result further supported the high reliability of the KIRIS open-response questions.

THE AMERICAN COLLEGE TESTING PROGRAM. The ACT is a college admissions testing program widely taken by high school students in Kentucky who plan on attending college. Kentucky sent ACT a data tape of grade 12 results for 1991-92, 1992-93 and 1993-94. Of the over 23,000 seniors who took the ACT in Kentucky in 1991-92, ACT was able to match ACT records to KIRIS results for some 18,000 students, providing comparisons in reading, mathematics and science. For 1993-94, an even greater match of more than 20,000 students was obtained.

The correlation between the ACT and KIRIS in 1991-92 was .48 for reading, .54 for mathematics, and .37 for science. The higher correlation for mathematics than for reading is consistent with the findings from the CTBS study. The lower correlation for science is also in line with expectations, since the KIRIS test is heavily weighted toward the measurement of science process, as directed by the Science Advisory Committee comprised of Kentucky teachers. For all three content areas, there was a clear trend that students rated higher by one testing program were also rated higher by the other. On the other hand, there were a substantial number of cases in which students with a high rating on one test received a low rating on the other. Again, this is in part a function of the limited number of open-response questions included in the KIRIS accountability tests in 1991-92, but also in part a function of the differences in the ways these two tests measure achievement.

For the 1993-94 data, the correlation between the ACT and KIRIS was .47 for reading, .67 for mathematics, and .46 for science. The correlations for mathematics and science are greater than in 1991-92, a logical result of increasing the test length for common items from 3 to 5 in changes made since the first year of the testing program. Conversely, the lack of a change in the correlation for reading is somewhat puzzling. Table 14-12 presents the correlations among selected ACT and KIRIS open-response tests in the lower diagonal, with disattenuated correlations in the upper diagonal, based on reliabilities on the diagonal. Reliability estimates for ACT tests are median correlations reported in a Preliminary Technical Manual.

Of the correlations between individual tests of the ACT and KIRIS, each KIRIS test has its highest correlation with the ACT test assessing the same content area. These correlations are, in turn, higher than correlations with other KIRIS tests, and generally lower than correlations among ACT tests. Interestingly, only one KIRIS test, mathematics, is more highly correlated with the ACT test assessing the same content area than with the composite ACT score. Disattenuated correlations yield fairly similar results. No appropriate comparison exists for KIRIS social studies.

TABLE 14-12
CORRELATION OF KIRIS OPEN-RESPONSE TEST RAW SCORES WITH ACT SCALE SCORES (n=20610)

Test		ACT				KIRIS				
		Reading	Math	Science	Composite	Reading	Math	Science	Social St.	Composite
ACT	Math	***.81***	.73	.93	1.03	.60	.61	.55	.56	.74
	Reading	.62	***.89***	.87	.91	.47	.82	.53	.48	.72
	Science	.74	.72	***.78***	1.03	.54	.73	.58	.54	.76
	Composite	.90	.84	.89	***.95***	.59	.73	.50	.54	.78
KIRIS	Reading	.47	.39	.42	.50	***.75***	.54	.57	.57	.80
	Math	.48	.67	.56	.62	.41	***.76***	.76	.55	.88
	Science	.43	.43	.46	.49	.38	.43	***.40***	.53	.78
	Social St.	.47	.42	.44	.51	.44	.43	.66	***.49***	.82
	Composite	.64	.66	.65	.74	.70	.74	.73	.85	***.94***

CHAPTER 15

CONSEQUENTIAL VALIDITY OF KIRIS AND THE KENTUCKY ASSESSMENT PROGRAM

CONSEQUENTIAL VALIDITY

An important aspect of an assessment's validity is whether it is used according to its intended purposes, or more broadly, whether its effects, or consequences, are aligned with its goals. This attention to "consequential validity" has been articulated most forcefully by Messick[1]. Messick links the *interpretation* of what a test represents or means (construct validity) with the *use* of a test or test information (consequential validity) as two essential aspects of validity. Consequential validity thus raises the question of the extent to which the Kentucky assessment program is achieving its goals and not having unintended, negative side effects.

INTENDED GOALS OF KENTUCKY ASSESSMENT PROGRAM

The role of the Kentucky assessment program, including the KIRIS assessments, is to promote educational improvement for all children in the state. It does this in three major ways:

1. The assessment program provides goals, standards, and criteria for educational achievement. These goals, standards, and criteria are found linked together throughout the assessment program. They include the statements of goals in the KERA legislation, the Academic Expectations, and the links between Academic Expectations and specific items and their scoring guidelines. The assessment program also includes operational definitions of success, various performance levels, relative weightings of assessment components, and so on.

2. The assessment program provides useful information on progress towards achieving those goals made by schools. Although the major informational use of KIRIS scores are in relation to school accountability as mandated in the KERA legislation, much assessment information is also provided that is useful for monitoring achievement and progress of individual students, the state as a whole, and various subgroups within the state.

3. The assessment program provides useful information on potential bias and differential impact of the assessment program in particular, and of education within the state, by region, and on various subgroups such as ethnic and racial minorities, children living in poverty, and so on.

EVIDENCE AND INTERPRETATION OF CONSEQUENTIAL VALIDITY

The KERA legislation and subsequent programs, including the assessment program, have engendered much discussion and activity. However, relatively little formal research is available on the impact of the assessment program on classroom practice, teacher development, or support of educational reform in Kentucky. In addition, in the available research it is often difficult to separate effects of the assessment program from other aspects of educational reform. For example, while the criteria for mathematics portfolios have made an impact on classroom instruction as a result of being included in the assessment system, some changes have resulted from the fact that Kentucky has adopted standards that effectively reflect the national standards promulgated by the National Council of Teachers of Mathematics.

---

[1] Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement (3rd Ed.)*. New York: ACE/MacMillan

It should be noted that the discussion in this chapter of the consequential validity of the Kentucky assessment program paints broad strokes, which may not apply to every classroom in every school. Change is taking place at different speeds and in different forms throughout the Commonwealth, often for different combinations of reasons. In addition, the assessment program, and people's understanding of it, has changed over time. Continuing research will be required to provide more complete results of the consequential validity of the Kentucky assessment program, as well as to keep the view up to date.

CONSEQUENCES: PROVIDES GOALS, STANDARDS, CRITERIA FOR INSTRUCTION AND CURRICULUM

Evidence continues to accrue regarding the effects of KIRIS on instructional practice, teachers' professional development, and support for educational reform.

IMPACT ON INSTRUCTIONAL PRACTICE

KIRIS has had a major impact on instructional practice, as evidenced by student products. Over 140,000 students in each of grades 4, 8, and 12 produced writing portfolios and mathematics portfolios, with teachers' assistance and direction. Relatively few portfolios were blank or incomplete, and the number of blanks and incompletes decreased from the initial year through subsequent years. For example, the Mathematics Portfolio Review Study found 5% of the Grade 8 portfolios were incomplete or blank in 1992-93, the first year, and 1% were incomplete or blank in 1993-94. To the extent that these portfolios represent worthwhile learning opportunities, KIRIS has had a major beneficial impact on instructional practice. In fact, support for the instructional value of portfolios is high, as indicated by surveys of teachers and principals. About two-thirds of those responding to a 1993-94 survey indicated "strong" or "very strong" agreement with portfolios as a learning approach.

Greater attention is being paid to student work, both for feedback to individual students and for improvement of school programs. In 1993-94 KDE provided student-level reports and an Interpretive Guide to schools. A report was provided for the parents/guardians of each student involved in accountability testing. Also, over 85% of schools reported taking advantage of the option to photocopy student work from the accountability tests. While the use of this information varies from school to school, it indicates that schools are willing to use the opportunities KIRIS provides to gather information about student performance for local uses. This use of accountability testing supplements the extensive use of Continuous Assessment in which schools are engaged. Over 65% of the schools in the state purchase Continuous Assessments, and 30% purchase Performance Assessment kits. These high percentages indicate that a considerable number of schools are incorporating assessments, and attending to instruction, that parallel the accountability tests.

Assertions of changes in classroom practice need to be substantiated by actual observations and other field-based research. The available field-based studies, while limited, agree that there are substantial changes in classroom practice and attitude. These changes are in the direction of being coherent with KIRIS and KERA. However, there are differences in sites, subject matter areas, and approaches.

Writing portfolios are the most studied subject matter area to date. One study, conducted by researchers based in the Appalachian Educational Laboratory (AEL), is focusing on education reform in four rural Kentucky districts since the

233

implementation of KERA in 1991-92. The AEL researchers have noted several changes associated with writing portfolios and writing instruction: "The major findings in the four rural school districts indicate:

"1. For most teachers, the state assessment program--which emphasizes portfolios and open-response questions--appeared to be the driving force behind their instructional changes.

"2. The major change at both accountable and non-accountable grade levels was an increased emphasis on writing and the writing process. At the 13 schools we studied, we saw and heard about writing activities that required students to think and create. Interviews with teachers, principals, and students indicated this emphasis on writing was relatively new, in most cases.

"3. Teachers had mixed reactions to the increased emphasis on writing. Generally, teachers who received the mosttraining on writing and portfolios were the most enthusiastic; however, many were unenthusiastic. The most common remarks we heard from teachers, listed in order of frequency, were these:

- Portfolios are time-consuming and burdensome.
- Students are burned out on writing and portfolios.
- Time spent on portfolios takes time away from teaching other skills.
- Students' writing and thinking skills have improved tremendously.
- Portfolios are an asset to the classroom and worth the extra effort.
- Portfolio standards are too high for most children to attain.

"4. Although the only across-the-board instructional change in the upper-elementary grades appeared to be the emphasis on writing, we saw and heard about a wide spectrum of other instructional strategies....

"5. Although KERA holds entire schools responsible for student achievement, we saw few effective efforts to plan and implement instructional changes on a school-wide basis. Such efforts were underway to varying degrees at most schools, but only one school appeared to have implemented a cohesive curriculum and instructional plan schoolwide."[2]As AEL cautioned, "these findings are based on work in four rural districts and cannot be generalized to the entire state."[3] However, it is useful to note the present state of research, and the model of longitudinal study that examines the relationship of assessment and instruction in context.

Pankratz and his colleagues[4], reporting preliminary results from data collected in 1994-95, found that significant changes in classroom practices were observed in classrooms where teachers had participated in extensive professional development, such as the professional development opportunities sponsored through PRISM (Partnership for Reform Initiatives in Science and Mathematics).

---

[2] Appalachian Educational Laboratory. (Dec., 1994). Instructional and Assessment in accountable and non-accountable grades, *Notes From The Field*, 4(1), 1-2.

[3] Ibid., p.4

[4] Pankratz, R., Ochs, D. et al. (April, 1995). *Configuration maps: Results from Kentucky*. Papers presented at the annual meeting of the American Educational Research Association, San Francisco, CA

IMPACT ON PROFESSIONAL DEVELOPMENT

It was recognized that teachers may need additional training and professional development to prepare students to reach the high goals reflected in the KIRIS assessments. Every indication shows that Kentucky teachers are availing themselves of extensive professional development, much of it substantively connected with the assessments. For example, over 300 educators have participated in the KERA Assessment Fellows program. Over 100 Distinguished Educators have been trained to help schools succeed in the assessment system as much as possible. Over 300 teachers participated in rescoring sessions during Summer, 1995 for writing and mathematics portfolios; evaluations strongly indicate that the participants found the activities very helpful to them in understanding the state standards and applying them to their classrooms. The fact that school districts paid for most of these volunteer teacher's participation also indicates the districts' and teachers' views of the professional value of the portfolio scoring sessions.

In 1993-94, teachers took advantage of the extensive professional development provided by KDE/ASME for Mathematics Portfolios:

- 14 KET telecasts, with an estimated viewing audience of 4,000-12,000 persons for each telecast;
- 2 full-day training workshops involving approximately 700 Mathematics Portfolio Cluster Leaders;
- 1 Optional Guided Scoring Practice Session, held regionally, involving one teacher from each school (approximately 1380 teachers throughout the state in 1993-94);
- 1 Summer Scoring Review Study, which involved over 100 teachers and resource specialists.

A report[5] conducted by a contractor for the U.S. Department of Education evaluating Chapter 1 technical assistance noted some influence of KERA. The study evaluated the activities and impact of Technical Assistance Centers (TACs) and Rural Technical Assistance Centers (RTACs) over three years on a select group of schools with programs funded by federal Chapter 1 funds that "had been identified as needing improvement on the basis of inadequate gains in student performance" (p. i). Of the nine sites, five were large urban districts throughout the United States; the single Kentucky site was a rural district. In Kentucky, "one of the most visible parts of the efforts to implement KERA mandates was the development of student assessments; teachers worked hard to prepare students for the new performance-based assessments... This was the only case [in the study] where we found evidence of TAC/RTACs working with districts to develop new assessments to fulfill Chapter 1 accountability requirements" (pp. 48-49).

A case study of mathematics portfolios is being conducted by Policy Study Associates, also funded by a contract by the federal government.[6] The study, now in its third year of data collection, involves a few schools in a single large district in Kentucky, as well as sites in four other states. The study intends to look at mathematics portfolios as a form of alternate assessment with the promise of improving learning and instruction. The study documents mixed feelings in the case study teachers, reflecting the challenges of having to "rethink both content and instruction," deal with time demands of implementing a new and complex program, wrestle with external accountability, and convince parents of the value of portfolio assessment and new approaches to mathematics instruction. Not surprisingly, the report finds implementation spotty and teacher attitudes mixed, but overall, "teachers were enthusiastic about the instructional possibilities of mathematics portfolios."

---

[5] Policy Studies Associates, Inc. (1994). *Third-year evaluation of the nine-site program initiative.* (A report to the U.S. Department of Education.) Washington, DC: Author.

[6] McCollum, H. et al. (August, 1994). *Portfolio assessment in mathematics: Views from the classroom. Annual report.* Washington, DC: Policy Studies Associates, Inc.

In related programs, in 1993-94 over 150 teachers P-5 participated in PRISM science workshops and Leadership Institutes in every region in the state. A central activity of these workshops is to develop curriculum units with embedded performance assessments that adhere to the state goals that reflect the KIRIS framework. Over 3200 teachers in 160 districts participated in writing portfolio analysis workshops. Over 1500 teachers have participated each year in Optional Guided Scoring Practice workshops for Writing Portfolios held regionally.

Neither KDE nor ASME monitors workshops and other activities regarding assessment not sponsored by KDE. Informal contacts, however, suggest that there is a high level of professional activity, both formal and informal, in helping teachers and administrators better prepare for the assessments. For example, T^2: Teachers to the Power of Two lists over 100 teachers willing to serve as resource persons dealing with assessment issues.

IMPACT ON SUPPORT FOR EDUCATIONAL REFORM IN KENTUCKY

The KIRIS assessment system and its role in accountability have generated controversy. Surveys find the assessment system--and the accountability it enacts--among the most highly debated areas in educational reform in Kentucky. For example, a survey conducted by Kentucky Institute for Educational Research (KIER) found that superintendents, parents, and business leaders supported the notion of school accountability and the KIRIS system. Principals gave mixed support. Teachers strongly did not support the current definition of school accountability nor the assessment system. It is not clear whether the KIRIS accountability system is having a positive or detrimental effect on support for educational reform in general. It is clear that accountability assessment is stressful to many people; it is yet to be seen whether it will engender cheating, short-sightedness, giving up hope, or other detrimental actions on a wide-spread or systematic basis. There is no evidence that such drawbacks are occurring now in any widespread way. Only about forty schools were reported to KDE for investigations in three years of testing, and fewer than half of these involved allegations of undue pressure, cheating by teachers or administrators on KIRIS, or unfair testing practices. The majority of these cases involved questions about the permissibility of testing accommodations for students with special needs.

The assessment system is controversial in large part because it makes clear assumptions and beliefs, many of which are different from those held by large segments of the population. These beliefs are not only about assessment, but about accountability, children's ability to learn, governmental role in education, the nature of knowledge, the relation of schooling to society, the nature of standards, and so on. KIRIS itself may do little to change these beliefs. However, it is clearly essential that KIRIS, in conjunction with the educational endeavor in general, must address these issues. The consequential validity of KIRIS is largely bound up with parents and other citizens' evaluation of education along these dimensions. The debate on accountability and assessment can be informed through continued, careful empirical research, but it will not be settled by judicious interpretation of data alone.

For example, the design of the KIRIS tests and the structure of the assessment system's rewards and sanctions assume that virtually "all children can learn at high levels." A study examined reactions to this assertion by Kentucky students, teachers, administrators, parents, school board members, and owners and managers of small businesses (Roberts & Kay, 1993)[7]. As summarized by Winograd et al. (submitted), the findings were:

---

[7] Roberts, R. & Kay, S. (September, 1993). *Kentuckians' expectations of children's learning: The significance for reform.* A public report prepared for the Prichard Committee for Academic Excellence and the Partnership of Kentucky School Reform, Lexington, KY: Roberts & Kay, Inc. (Available from the Prichard Committee for Academic Excellence, P.O. Box 1658, Lexington, KY 40592-9980.)

"1. The present structure of schools and of society profoundly affects what Kentuckians consider possible for children. Many of the study's participants believe that the stratification in the schools reflects God-given differences in ability and mirrors the stratification in society. One administrator said, "What's a high level for someone may not be for somebody else. I mean, there have got to be people who work at McDonald's, and if everybody is a genius, who is going to be the genius who is going to want to work at McDonald's?"

The less "schooled" children are, the more participants believe in their capacity. Most of the participants had the highest expectations of children from birth to the end of the primary school. Teachers and students were more likely than the other participants to disagree with the assertion that all children can learn at high levels. Most participants defined learning at high levels in a traditional manner: being quick at gaining knowledge that will result in high test scores and good grades.

2. Most participants view assigning blame for children's and schools' failures as important responsibilities; the question of blame dominates conversations about schooling, forcing out attention to needed changes. Teachers, parents, citizens, and business people blame parents for most of the failures at school. When participants consider their own school failures, they do not blame their own parents. Instead, they blame individual educators. Students place some blame on parents, but more blame students themselves" (pp. 9-10).[8]

A study sponsored by the National Science Foundation (NSF) found that Kentucky's assessment system met three criteria that indicated the assessment system could contribute to systemic educational reform in science and mathematics. Policy Studies Associates (PSA)[9] found that Kentucky's assessment system was one of only a few state-level assessment systems that met the criteria of: 1) the assessment system was aligned with the state's science and mathematics reform objectives; 2) the assessment included the grade levels where the NSF-sponsored work was being carried out (a minority of states have a state assessment that includes science, and even fewer have science assessments at multiple grade levels, as does Kentucky); 3) the assessment system was or will be operational in time to provide useful information, including baselines, during the period of science and/or mathematics reform where NSF was a partner with the state. According to the PSA analysis, Kentucky was one of seven states that met the three criteria in math, and one of only five in science. Figure 15-1, based on similar figures in the PSA report, displays names of states meeting these criteria. State name abbreviations are placed in parentheses to indicate states partially meeting one or more criteria.

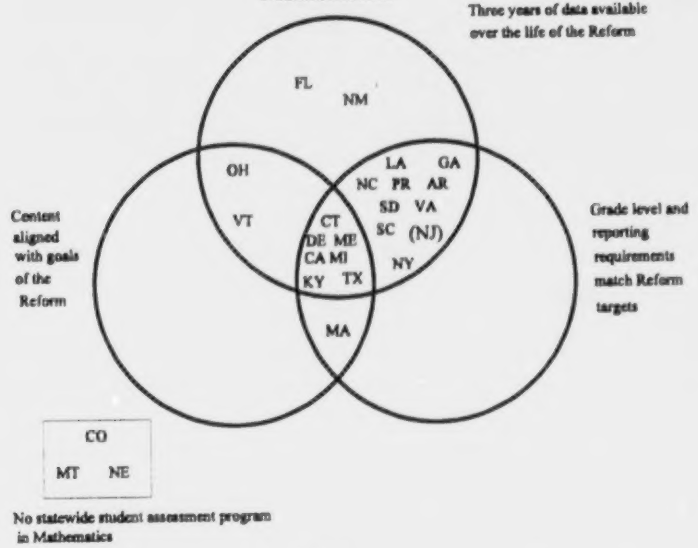
---

[8] Winograd, P., Jones, D., & Perkins, F. (submitted). *The politics of alternative assessment: Lessons from Kentucky*. (Manuscript obtained from first author.)

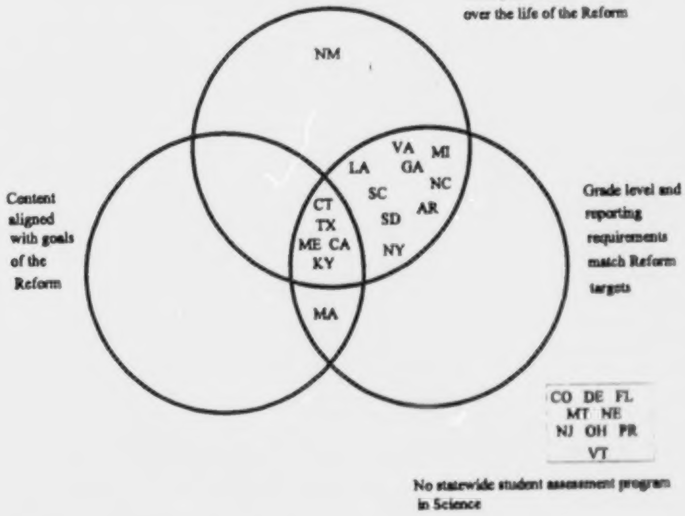
[9] Laguarda, K. G., Breckenridge, J. S., Hightower, A.M., & Adelman, N. E. (September,1994). *Assessment programs in the statewide systemic initiatives (SSI) International, primary contractor.)* Prepared under contract for the National Science Foundation, SRI International, primary contractor. Washington, DC: Policy Studies Associates, Inc.

FIGURE 15-1. STATES WITH NSF-SPONSORED SYSTEMIC REFORM INITIATIVES IN MATHEMATICS AND SCIENCE WHOSE ASSESSMENT SYSTEMS MEET CERTAIN CRITERIA.

Mathematics


Three years of data available over the life of the Reform
Content aligned with goals of the Reform
Grade level and reporting requirements match Reform targets
FL
NM
OH
VT
LA
GA
NC
PR
AR
SD
VA
SC
(NJ)
NY
CT
DE
ME
CA
MI
KY
TX
MA
CO
MT
NE
No statewide student assessment program in Mathematics

Science


Three years of data available over the life of the Reform
Content aligned with goals of the Reform
Grade level and reporting requirements match Reform targets
NM
VA
MI
LA
GA
NC
SC
AR
SD
NY
CT
TX
ME
CA
KY
MA
CO
DE
FL
MT
NE
NJ
OH
PR
VT
No statewide student assessment program in Science

BEST COPY AVAILABLE

CONSEQUENCES: PROVIDES INFORMATION ON STATUS AND PROGRESS

KIRIS provides information to schools and districts in several forms. These are described more fully in the chapter on Reporting in this Technical Manual. Essentially, KIRIS reports include scores on each subject matter area and a non-cognitive index for schools and districts each year. KDE/ASME produce a biennium report that summarizes each school's performance in terms of the baseline and two subsequent years of the first accountability cycle, and summarizes the school's status in relation to rewards and assistance. Student reports are produced each year for subject matter areas, drawing on the common items and writing portfolio, and excluding the matrix-sampled items. The student reports are sent by KDE/ASME to schools, where schools decide how to distribute them.

In addition, each year KDE/ASME produces a summary report for the state, by region, by gender and by ethnic group. KDE/ASME have also issued reports relating KIRIS scores to NAEP scales as those data have been available in reading and mathematics. These reports provide a national comparison for those subject matter areas.

Schools, districts, and classroom teachers have reported using the score reports in a variety of ways consistent with the intent of KIRIS. The most common use is in broad program review to mark progress over a year or two, and focus resources for instructional program improvement. However, KIRIS scores have been used by schools and teachers for other purposes. There are increasing requests to have KIRIS provide information in addition to the school accountability function it was originally designed to provide. There have been wide-spread calls for additional information in the other traditional evaluation areas:[10]

1. for example, individual student achievement status, for use on report cards;

2. individual student comparative status, for monitoring Kentucky achievement in relation to national or regional achievement, or for college admissions;

3. school comparisons (the news media routinely convert KDE/ASME reports into "rankings" that facilitate comparisons between schools and districts);

4. student diagnostic information for monitoring student progress and informing instructional changes by classroom teachers;

5. instructional program evaluation, to monitor and improve instructional programs, school curricula, and inform teacher assignments and professional development.

While KIRIS results are currently provided that address these areas to some extent, most would require substantial changes in design and or operation. Some of these uses are possible currently; some are possible but somewhat incompatible with the intended uses of KIRIS results. For example, using student KIRIS results as the sole basis for school report card grades is viewed by KDE as an inappropriate use of KIRIS results. Providing diagnostic information for individual students would require much more rapid feedback than is currently possible, as well as local interpretation. KDE holds that such diagnosis is more appropriately undertaken by classroom teachers using classroom assessments other than, or at least in addition to, KIRIS assessments.

---

[10] For example, see the report done by The Evaluation Center, Western Michigan University, (January, 1995). *An independent evaluation of the Kentucky Instructional Results Information System (KIRIS).*

CONSEQUENCES: IS FAIR TO ALL

It is important that the Kentucky accountability program provide a fair challenge to all schools. This is especially true regarding the consequences of rewards and assistance based on the assessment results. Several factors have been examined to explore whether the Kentucky accountability program is fair to schools. The factors include geographical location of school, racial/ethnic composition, economic status of students, initial baseline score, school size, and grade level organization. Based on these analyses, the Kentucky assessment program appears to be fair in that rewards and assistance were distributed across these dimensions without statistically significant unevenness. The exception was grade level, where proportionally more elementary schools received rewards than did middle schools or high schools. This result appears to be explained by differences in education at these grade levels, and does not necessarily represent bias. Several analyses are summarized below.

DESCRIPTION OF VARIABLES

SCHOOL RESULT (ACCOUNTABILITY PERFORMANCE JUDGEMENT) - Each school was assigned a reward/assistance category, based on the results of the KIRIS scores.

- Decline - school improvement plan required, assigned Distinguished Educator for assistance;
- Improving (school improvement plan required);
- Successful-Year 2 (did not meet improvement goal based on the average of scores from both years of the second biennium, but did reach improvement goal in Year 2 of the biennium, 1993-94; no school improvement plan required);
- Successful (no requirements);
- Reward (financial reward given).
- Uncategorized (no judgment made regarding Reward/Assistance Status at the request of the school, because the school did not exist at the beginning of the accountability cycle, because data was lost due to fire, or because the school was very small).

These reward/assistance categorizations were recorded under the variable "SCHOOL RESULT." For purposes of regressions and correlations, these reward/assistance categories were transformed into a number ordered, from most additional requirements on school to least, with Decline being assigned a 1, Improving a 2, and so on, with Uncategorized schools not being included in the analysis.

BASELINE - Baseline score from 1991-92.

INDEX - 93 - Overall school index for 1992-93.

INDEX - 94 - Overall school index for 1993-94.

OVERALL INDEX - Average of Index-93 and Index-94 for the school.

PCTGOAL - The variable, Percent of Improvement Goal, represents how much the school achieved in terms of improving scores to its required improvement goal. Percent of Improvement Goal was calculated as the ratio of gain, the difference between Overall Index and Baseline, to the difference between a school's improvement goal and its baseline. PCTGOAL is a continuous variable related to the categorical reward/assistance status.

ETHNICITY (Proportion Racial/Ethnic Minority) - Based on data reported to KDE by schools, the numbers of students in each school were identified in terms of major ethnic/racial groups: Caucasian, Hispanic, African-American, Asian, and Native American. The percentage of students who were not Caucasian was calculated.

WEALTH (Economic Status of Students) - Using data collected by School and Community Nutrition within the Kentucky Department of Education, the percentage of students in each school receiving free or reduced-price school lunch was calculated. This variable indicates the economic status of students' households, with lower values indicating greater community wealth.

ENROLLMENT(Accountability Grade Enrollment) - The total number of students enrolled in the accountability grade (i.e., Grade 4, 8, or 12, or 4 and 8 combined, or grade 8 and 12 combined for schools having two accountability grades) was obtained. Data were not available on class size.

ACCGRADE (Accountability Grade) - The accountability grade(s) in the school. The grades were, Grade 4, 8, 12, or 4 and 8 combined, or 8 and 12 combined. For the analyses, the values used were 4, 6 (Grades 4 & 8 combined), 8, 10 (Grades 8 & 12 combined), and 12.

PREDICTABILITY OF SCHOOL ACHIEVEMENT

The question may be asked, "What is the relation of school achievement across the state to variables such as economic status of the students' households, percent of minority students in the school, and so on?" Another way of phrasing this question is, "Are schools equally likely to achieve, no matter what size the schools are, where they started in terms of baseline score, and so on?" This concern with equitable impact of KIRIS is a central point of investigation following the conclusion of the first accountability cycle.

The question may be addressed by using a multiple regression equation to describe the statistical relationships between percent of Improvement Goal obtained and five variables: Accountability Grade reflecting an elementary, middle, high school, or combined elementary-middle or middle-high school (ACCGRADE); initial Baseline score in 1992, reflecting where the school started (BASELINE); percentage of minority students in the school (ETHNICITY); percentage of students receiving free or reduced school lunch (WEALTH); and, size of school in terms of student enrollment at the accountability grade (ENROLLMENT).

Percent of Improvement Goal obtained by schools throughout the state was not strongly predicted by any variables, taken singly or in combination. The "best fit" multiple regression including the five variables noted above had an R-square of approximately .17. This can be interpreted as meaning that about 17% of the variance in Percent of Improvement Goal could be accounted for by these six variables, and 83% of the variance was due to other factors. The complete correlation matrix for all the variables used in the differential impact analyses, for all schools, is given in table 15-1. Percent of Improvement Goal and School Result are highly correlated, as would be expected. Similarly, School Result is somewhat less well predicted (R-square = .14) than Percent of Improvement Goal by the five variables used in this analysis.

A similar question can be raised concerning region: "Are schools equally likely to achieve, no matter the region in which schools are located?" For this analysis region was coded as seven 1-0 dummy variables representing the eight regions of Kentucky, and used to predict Percent of Improvement Goal. The "best fit" multiple regression for Region resulted in an R-square of approximately 0.01. Thus only one percent of the variance in Percent of Improvement Goal could be accounted for by region.

TABLE 15-1
SCHOOL-LEVEL CORRELATIONS OF DEMOGRAPHIC VARIABLES AND ACCOUNTABILITY INDICES (N=1247* SCHOOLS)

NAME	BASELINE	INDEX 93	INDEX 94	OVERALL INDEX	PCTGOAL	SCHOOL RESULT	ETHNICITY	WEALTH	ENROLLMENT	ACCGRADE
BASELINE	1.00	.65	.59	.69	-.20	-.17	-.06	-.68	.35	.42
INDEX_93	.65	1.00	.59	.88	.45	.35	-.16	-.48	.08	.04
INDEX_94	.59	.59	1.00	.90	.55	.49	-.13	-.45	.10	.12
OVERALL INDEX	.69	.88	.90	1.00	.56	.48	-.16	-.52	.11	.10
PCTGOAL	-.20	.45	.55	.56	1.00	.84	-.13	.06	-.24	-.36
SCHOOL RESULT	-.17	.35	.49	.48	.84	1.00	-.12	.02	-.20	-.31
ETHNICITY	-.06	-.16	-.13	.16	-.13	-.12	1.00	.08	.14	-.04
WEALTH	-.68	-.48	-.45	-.52	.06	.02	.08	1.00	-.40	-.40
ENROLLMENT	.35	.08	.10	.11	-.24	-.20	.14	-.40	1.00	.56
ACCGRADE	.42	.04	.12	.10	-.36	-.31	-.04	-.40	.56	1.00

*The number of schools is slightly lower for correlations involving variables with missing data; n = 1243 for School Result, n = 1222 for Wealth, and n = 1208 for Ethnicity.

RELATIONSHIP BETWEEN ACCOUNTABILITY PERFORMANCE JUDGEMENT AND ACCOUNTABILITY GRADE LEVEL

Grade level is strongly related to obtaining rewards or assistance in the Kentucky accountability program. As documented in table 15-2, proportionally more elementary schools (those with Grade 4) achieved reward status than did middle or high schools. Of the variables included in the regression analysis presented above, Accountability Grade was the single demographic variable mostly strongly correlated with School Result (r = -.31). Accountability Grade by itself accounted for essentially half of the variance in the multiple regression reported above.

Plausible reasons for these grade-level effects include factors of instructional programs and school organization, and so do not inherently appear unfair. For example, some people have noted that it is usually easier for smaller schools than larger to make changes in curriculum and instruction, and for elementary schools to make such changes than high schools. Others have noted that changes in instruction may be offset by previous instructional experiences of students, and that the three years since the passage of KERA represents a proportionally large amount of elementary students' school experience, while it represents a much smaller proportion of high school students' experience. Other observers have noted the differences in personality and professional training of elementary, middle, and high school teachers, and speculated on the role this may play in fitting into changes such as those mandated by KERA and evaluated by KIRIS. This phenomenon of grade-level differences is under further study by KDE. To this end, the regression analyses undertaken above and several individual correlations were re-analyzed by Accountability Grade. Although results of several analyses by Accountability Grade are presented below with appropriate emphasis on issues of sample size, it is worth noting in advance that the small number of combined Grade 8 and 12 Schools (a maximum of 36, when some schools with missing data are included in the analysis) renders regression results questionable at best when more than one variable is used for prediction, i.e., for any regression analysis other than a correlation.

PREDICTABILITY OF SCHOOL ACCOUNTABILITY PERFORMANCE JUDGEMENT AT EACH ACCOUNTABILITY GRADE LEVEL

Two regression analyses, parallel to those described above, were performed for schools having each of the five Accountability Grade classifications. The maximum number of schools in each classification is given in table 15-2; actual numbers were somewhat smaller in analyses involving Ethnicity and/or Wealth. Results of these regression analyses are presented in table 15-3.

TABLE 15-2
REWARD/ASSISTANCE CLASSIFICATION BY GRADE LEVEL

Classification	Grade 4		4 & 8		8		8 & 12		12		Grand Total	
Categories	N	%	N	%	N	%	N	%	N	%	N	%
Reward	345	50.9	51	40.2	56	27.2	9	25.0	24	12.0	485	38.9
Success	51	7.5	12	9.4	20	9.7	2	5.6	12	6.0	97	7.8
Success Yr. 2	112	16.5	25	19.7	45	21.8	7	19.4	80	40.0	259	21.6
Improving	154	22.7	36	28.3	78	37.9	15	41.7	55	27.5	338	27.1
Decline	13	1.9	2	1.6	7	3.4	3	8.3	29	14.5	54	4.3
Uncategorized	3	0.4	1	0.8	0	0.0	0	0.0	0	0.0	4	0.3
Total	678	100.0	127	100.0	206	100.0	36	100.0	200	100.0	1247	100.0

TABLE 15-3
PERCENT OF VARIANCE EXPLAINED IN PERCENT OF IMPROVEMENT GOAL, BY ACCOUNTABILITY GRADE*

Variables Predicting Percent of Improvement Goal	Grade 4	Grade 4 & 8	Grade 8	Grade 8 & 12	Grade 12	All Schools
Baseline, Ethnicity, Wealth, and Enrollment	.08 (.08)	.06 (.02)	.08 (.06)	.31 (.21)	.12 (.10)	.10 (.10)
Region Alone	.03 (.02)	.12 (.07)	.03 (.00)	.12 (-.07)	.03 (-.01)	.01 (.01)
Baseline, Ethnicity, Wealth, Enrollment, and Region	.09 (.08)	.21 (.13)	.11 (.05)	.38 (.09)	.13 (.08)	.11 (.10)

*R-square, adjusted for shrinkage is given in parentheses.

From these analyses, the only substantial departures from results obtained for all schools occur for the Accountability Grade classifications with the smallest numbers of schools. The extremely small number of combined Grade 8 and 12 schools (n=36 when region alone is entered into the regression analysis, n = 33 when the other variables are considered) renders the results obtained practically uninterpretable; some question exists, only to a much lesser extent, for combined Grade 4 and 8 schools. When r-square is adjusted for shrinkage based on the number of schools and the number of variables included in the analysis, it drops appreciably for the combined Grade schools. Indeed, only the Grade 8 and 12 regression of Baseline, Percent Ethnicity, Percent Wealth and Accountability Grade Enrollment yields an adjusted r-square (.21) substantially higher than that obtained for all schools. In this case, accounting for twenty-one percent of the variance leaves nearly eighty percent of the variance to be explained by other variables.

RELATIONSHIP BETWEEN SCHOOL ACCOUNTABILITY PERFORMANCE JUDGEMENT AND BASELINE SCORE

For all schools in the state taken together, Baseline scores had a slight negative relationship with eventual Reward/Assistance status (r = -.17). However, when schools are considered by accountability grade, as shown in Table 15-4, there is a positive correlation in Grade 8 and Grade 8 and 12 schools. Note that the Grade 4, Grade 8, and Grade 12 and 12 schools' correlations are significantly different from zero at the .05 level of statistical significance. Given that the result for Grade 8 and 12 schools should be discounted to a large extent because of the very small number of schools upon which it is based, it can be concluded that where correlations are statistically different from zero, they are negative and, more importantly, small in magnitude. A correlation of -.23 at grade 12 means that baseline only accounts for about five percent of the variance in percent of improvement goal obtained.

TABLE 15-4 CORRELATION OF BASELINE AND PERCENT OF IMPROVEMENT GOAL, BY GRADES						
Category	Grade 4	Grade 4&8	Grade 8	Grade 8 & 12	Grade 12	All Schools
Correlation	-.08	-.05	.13	.35	-.23	-.20
Probability	.03	.58	.06	.04	.01	.001

This finding is important because it informs the discussion of the appropriateness of the underlying model of determining "successful" schools. Some people have argued that the lower the baseline, the more difficult it would be for a school to achieve reward status. This argument was based on the fact that the lower the baseline, the greater the number of points a school must improve to meet its improvement goal; also, some people pointed to the difficulty historically low-achieving schools have struggled with in raising student achievement. Other people have argued that a higher baseline score would be more difficult to improve on, perhaps reflecting the opinion that fostering learning necessary to improve student performance from Apprentice to Proficient is a more difficult task than fostering learning necessary to improve student performance from Novice to Apprentice. The findings from this first accountability cycle support the general view that schools with initially lower scores may have a slightly easier time making the progress necessary to achieving reward status.

RELATIONSHIP BETWEEN SCHOOL ACCOUNTABILITY PERFORMANCE JUDGEMENT AND STUDENT ECONOMIC STATUS

For all schools in the state taken together, Wealth had a slight positive relationship with Percent of Improvement Goal (r - .06), a relationship indicating that schools with a higher percentage of students eligible for free/or reduced lunch were somewhat more likely to improve in their scores than schools with a lower percentage. Considered by grade, however, the correlations are negative, and the Grade 8 and Grade 8 and 12 schools' correlations are significantly different from zero at the .05 level of statistical significance. Given that the result for Grade 8 and 12 schools should be discounted to a large extent because of the very small number of schools upon which it is based, it can again be concluded that where correlations are statistically different from zero, they are negative and, more importantly, small in magnitude. A correlation of -.26 at grade 8 means that percent free lunch only accounts for about seven percent of the variance in Percent of Improvement Goal.

TABLE 15-5 CORRELATION OF PERCENT FREE/REDUCED LUNCH AND PERCENT OF IMPROVEMENT GOAL, BY GRADE						
Category	Grade 4	Grade 4&8	Grade 8	Grade 8 & 12	Grade 12	All Schools
Correlation	-.06	-.08	-.26	-.53	-.11	.06
Probability	.11	.35	.001	.001	.13	.04

It is recognized that participation in free and reduced price school lunch programs is not the same as a direct measure of students' household income. Nonetheless, KIRIS assessments appear fair to schools with many students from low-income households.

The school level findings were corroborated by an independent analysis that compared schools' reward status with the districts' assessed property value, done by *The Lexington Herald-Leader*[11], one of the largest newspapers in Kentucky. The analysis ranked districts by KIRIS test achievement and by wealth. KIRIS test achievement was defined as the percent of the improvement goal achieved in the 1992-94 final accountability index (same as PCTGOAL used in other analyses reported here but for districts rather than individual schools). This "percent of goal" ranged from 131.4% to 88.8%. The districts were then ranked, from highest percent achieved to lowest.

Wealth was defined as the district's amount of assessed (taxable) property value per student enrolled in the district. This "wealth" ranged from $483,082 to $51,479 per student. Districts were also ranked by this wealth index. The *Herald-Leader* analysis found that "poor districts were just as likely as rich ones to get cash rewards for improvement on tests." As evidence the newspaper observed that "Eleven of the 20 highest scoring school districts and half of the 42 getting rewards rank in the bottom half of all districts in property wealth." The newspaper also noted that the school that improved the most was in a county among the 15 poorest in the state. The full results are reproduced in Table 15-6.

Over the three years of the 1991-94 accountability cycle, the progress schools made was not related to their students' participation in free/reduced price school lunch. A school with a high proportion of students in economically distressed status was just as likely to achieve a reward/assistance level as a school with relatively fewer poor students. Note that this result does not imply that schools achieved at the same absolute score levels--only that the model of measuring schools against their own baselines did not discriminate for or against schools with high proportions of students from households in poverty. The relationship between percent of students receiving free/reduced-price school lunches and school achievement decreased dramatically over the three years of KIRIS assessments. The very strong correlation between these two variables -- initially $r = -.68$ -- dropped to $r = -.52$ for the average index for 1993-94. Taken in conjunction with the other information just presented, this is a crucial result, since the original lawsuit that resulted in KERA was brought on the basis of inequitable funding. Other studies[12] have shown that the funding gap has lessened since the passage of KERA. Figure 15-2 displays the changes in per-pupil funding over the past five years, comparing the richest and poorest twenty percent of all districts. Results indicate a marked change toward equalization of funding in the first two years, followed by relatively little change in the past two years.

---

[11] *Lexington Herald-Leader*, Feb. 9, 1995, pp. A1, A10. "Poor schools as likely as rich to hit KERA goals, earn rewards."

[12] Goetz, S. J. & Debertin, D. L.. (1994). School finance reform. In *A review of research on the Kentucky Education Reform Act (KERA)*. Report prepared by the University of Kentucky/University of Louisville Joint Center for the Study of Educational Policy, for the Kentucky Institute for Education Research (KIER). Frankfort, KY:KIER.

TABLE 15-6
RANKINGS BY DISTRICT ACHIEVEMENT AND DISTRICT WEALTH. (FROM THE *HERALD-LEADER*)

	% of Goal	Rank by Goal	Rank by Wealth	District	% of Goal	Rank by Goal	Rank by Wealth
Science Hill Ind.	131.4%	1	163	Bell Co.	104.0%	27	160
Williamsburg Ind.	118.7%	2	134	Henry Co.	103.8%	28	69
Southgate Ind.	116.1%	3	12	Somerset Ind.	103.4%	29	35
Morgan Co.	111.8%	4	158	Eminence Ind.	103.4%	30	74
Campbellsville Ind.	109.5%	5	85	Erlanger-Elsmere	103.3%	31	18
Murray Ind.	109.2%	6	22	Pikeville Ind.	103.3%	32	45
Harrison Co.	108.9%	7	93	Todd Co.	103.3%	33	108
Oldham Co.	108.5%	8	23	Raceland Ind.	102.8%	34	84
Ludlow Ind.	108.3%	9	148	Trimble Co.	102.7%	35	26
Walton Verona Ind.	108.2%	10	119	Fayette Co.	102.7%	36	2
Johnson Co.	108.2%	11	166	Elizabethtown Ind.	102.7%	37	53
Glasgow Ind.	107.4%	12	57	Mayfield Ind.	102.6%	38	72
Calloway Co.	107.1%	13	48	Bracken Co.	102.4%	39	94
Magoffin Co.	106.7%	14	170	Burgin Ind.	102.4%	40	30
Rockcastle Co.	106.7%	15	152	Carlisle Co.	102.3%	41	92
Barbourville Ind.	106.4%	16	146	Harrodsburg Ind.	102.3%	42	61
Ft. Thomas Ind.	106.3%	17	19	Mercer Co.	102.2%	43	49
Anchorage Ind.	106.3%	18	1	Lawrence Co.	102.1%	44	132
Whitley Co.	106.3%	19	149	Pineville Ind.	102.1%	45	141
Jenkins Ind.	105.8%	20	174	Marshall Co.	101.9%	46	25
Pendleton Co.	105.6%	21	111	Grayson Co.	101.8%	47	103
Dawson Springs In.	105.3%	22	165	Owen Co.	101.7%	48	105
Muhlenburg Co.	104.6%	23	76	Hazard Ind.	101.5%	49	113
Lee Co.	104.4%	24	150	Adair Co.	101.5%	50	124
Elliott Co.	104.3%	25	164	Paintsville Ind.	101.2%	51	80
Edmonson Co.	104.0%	26	145	Caverna Ind.	101.2%	52	100

TABLE 15-6
RANKINGS BY DISTRICT ACHIEVEMENT AND DISTRICT WEALTH. (FROM THE *HERALD-LEADER*)

District	% of Goal	Rank by Goal	Rank by Wealth	District	% of Goal	Rank by Goal	Rank by Wealth
Beechwood Ind.	101.1%	53	4	Cumberland Co.	99.1%	79	107
Warren Co.	100.9%	54	38	East Bernstadt Ind.	99.0%	80	176
Powell Co.	100.7%	55	162	Spencer co.	99.0%	81	95
Rowan Co.	100.7%	56	89	Scott Co.	98.0%	82	27
Washington Co.	100.7%	57	67	Allen Co.	98.8%	83	116
Boyle Co.	100.7%	58	29	Boone Co.	98.7%	84	3
Perry Co.	100.5%	59	129	Meade Co.	98.6%	85	126
Cloverport Ind.	100.5%	60	51	Breckinridge Co.	98.6%	86	78
Hopkins Co.	100.5%	61	51	Metcalfe Co.	98.5%	87	112
Henderson Co.	100.5%	62	41	Corbin Ind.	98.5%	88	98
Monroe Co.	100.2%	63	133	Graves Co.	98.4%	89	77
Laurel Co.	100.2%	64	104	Monticello Ind.	98.4%	90	172
Trigg Co.	100.2%	65	43	Kenton Co.	98.3%	91	9
Clark Co.	100.2%	66	39	Menifee Co.	98.2%	92	137
Anderson Co.	100.2%	67	28	Owensboro Ind.	98.2%	93	31
Daviess Co.	100.0%	68	37	Robertson Co.	98.0%	94	101
Hancock Co.	99.8%	69	44	Russell Co.	97.9%	95	88
Nelson Co.	99.8%	70	52	Greenup Co.	97.7%	96	121
McCreary Co.	99.7%	71	175	Lyon Co.	97.5%	97	11
Woodford Co.	99.6%	72	6	Mason Co.	97.4%	98	16
Marion Co.	99.5%	73	75	Jackson Co.	97.3%	99	168
Russellville Ind.	99.4%	74	122	Taylor Co.	97.3%	100	79
Hickman Co.	99.3%	75	71	Floyd Co.	97.2%	101	135
Butler Co.	99.3%	76	138	Lewis Co.	97.2%	102	143
Union Co.	99.3%	77	54	Carroll Co.	97.2%	103	20
Pike Co.	99.2%	78	115	Franklin Co.	97.1%	104	91

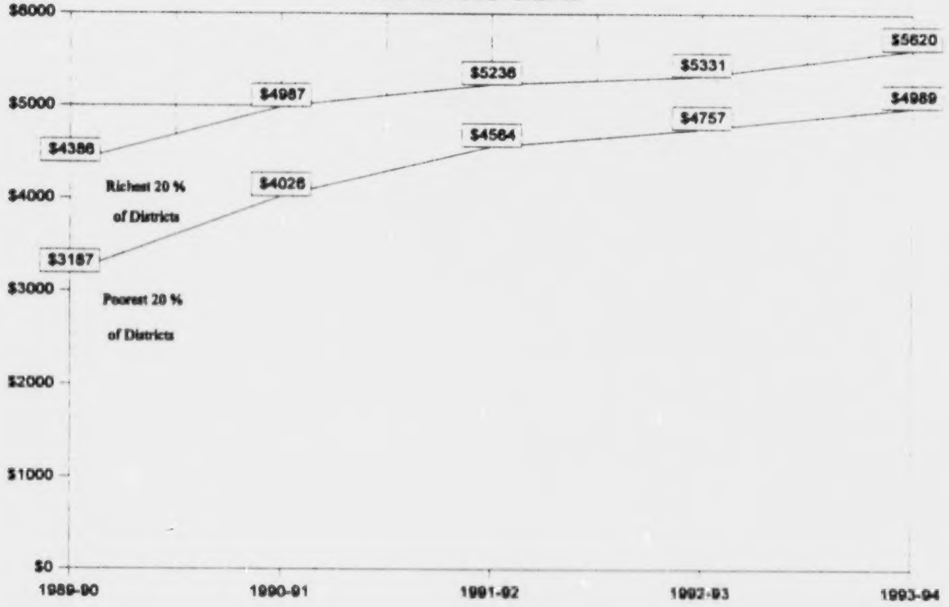
TABLE 15-6
RANKINGS BY DISTRICT ACHIEVEMENT AND DISTRICT WEALTH.
(FROM THE *HERALD-LEADER*)

District	% of Goal	Rank by Goal	Rank by Wealth	District	% of Goal	Rank by Goal	Rank by Wealth
Green Co.	97.1%	105	99	Lincoln Co.	95.4%	131	128
Russell Ind.	97.1%	106	32	Fleming Co.	95.1%	132	96
Ohio Co.	97.0%	107	83	McCracken Co.	95.0%	133	40
Franklin Co.	96.8%	108	13	Martin Co.	95.0%	134	109
Garrard Co.	96.8%	109	50	Casey Co.	94.9%	135	123
Bourbon Co.	96.7%	110	47	Bath Co.	94.7%	136	136
Livingston Co.	96.7%	111	24	Knox Co.	94.6%	137	153
Providence Ind.	96.6%	112	139	Jessamine Co.	94.5%	138	34
Pulaski Co.	96.6%	113	70	Webster Co.	94.5%	139	33
Madison Co.	96.5%	114	56	Logan Co.	94.3%	140	60
Caldwell Co.	96.4%	115	73	Bowling Green Ind.	94.2%	141	36
Jefferson Co.	96.2%	116	5	Montgomery Co.	94.2%	142	102
Ballard Co.	96.2%	117	66	Hart Co.	94.1%	143	127
Carter Co.	96.2%	118	151	Jackson Ind.	94.1%	144	159
Clinton Co.	96.1%	119	97	McLean Co.	94.1%	145	58
Knott Co.	96.1%	120	142	Bullitt Co.	94.0%	146	65
Paris Ind.	96.0%	121	82	Christian Co.	94.0%	147	87
Shelby Co.	96.0%	122	14	Wayne Co.	93.8%	148	140
Williamston Ind.	95.9%	123	62	Danville Ind.	93.7%	149	15
Hardin Co.	95.9%	124	62	Grant Co.	93.6%	150	106
Campbell Co.	95.9%	125	8	Paducah Ind.	93.5%	151	21
Crittenden Co.	95.8%	126	86	Middlesboro Ind.	93.0%	152	81
Clay Co.	95.7%	127	157	Silver Grove Ind.	92.8%	153	117
Ashland Ind.	95.7%	128	42	West Point Ind.	92.7%	154	161
Berea Ind.	95.6%	129	120	Owsley Co.	92.7%	155	169
Leslie Co.	95.4%	130	118	Frankfort Ind.	92.7%	156	7

TABLE 15-6
RANKINGS BY DISTRICT ACHIEVEMENT AND DISTRICT WEALTH.
(FROM THE *HERALD-LEADER*)

District	% of Goal	Rank by Goal	Rank by Wealth
Wolfe Co.	92.2%	157	167
Harlan Co.	92.1%	158	156
Harlan Ind.	92.0%	159	147
Simpson Co.	92.0%	160	46
Gallatin Co.	91.9%	161	63
Estill Co.	91.8%	162	155
Barren Co.	91.7%	163	55
Boyd Co.	91.6%	164	17
Newport Ind.	91.5%	165	130
Fairview Ind.	91.5%	166	131
Fulton Co.	91.4%	167	90
Augusta Ind.	91.4%	168	125
Letcher co.	91.2%	169	154
Covington Ind.	91.1%	170	68
Bellevue Ind.	90.2%	171	59
Dayton Ind.	90.1%	172	171
LaRue Co.	89.9%	173	110
Nicholas Co.	89.8%	174	114
Bardstown Ind.	89.8%	175	10
Breathitt Co.	88.8%	176	144

FIGURE 15-2
PER-PUPIL REVENUE,
SEEK SCHOOL FUNDING



Source: Kentucky Department of Education
Office of District Support Services,
Division of Finance

RELATIONSHIP BETWEEN SCHOOL ACCOUNTABILITY PERFORMANCE JUDGEMENT AND RACIAL/ETHNIC MINORITY ENROLLMENT

For all schools in the state taken together, having higher proportions of minority students was related to lower school performance (r =-.13). When schools are considered by accountability grade, as shown in Table 15-7, there actually is a very small, positive correlation in Grade 12 schools, and negative or zero correlations at other Grades. Most notably, however, only the Grade 4 school correlation is significantly different from zero at the .05 level of statistical significance, and it is small in magnitude. A correlation of -.22 at grade 4 means that percent of minority enrollment only accounts for about five percent of the variance in percent of threshold obtained.

TABLE 15-7
CORRELATION OF RACIAL/ETHNIC MINORITY ENROLLMENT AND PERCENT OF IMPROVEMENT GOAL, BY GRADE

Category	Grade 4	Grade 4 & 8	Grade 8	Grade 8 & 12	Grade 12	All Schools
Correlation	-.22	.003	-.08	-.12	.03	-.13
Probabilility	.001	.98	.26	.49	.64	.001

It should be noted that more than half of the schools in Kentucky serve student bodies where 9% or fewer of the students are minority. African-American students--the only sizable racial minority group in Kentucky schools--have scored lower than Caucasian students on the KIRIS tests, as reported elsewhere in this *Technical Manual*. Understanding these differences and informing appropriate action will require additional investigation.

CHAPTER 16
GENDER AND ETHNIC DIFFERENCES

PERFORMANCE OF SUBGROUPS ON MULTIPLE-CHOICE AND OPEN-ENDED ITEMS

A crucial question in any testing program is that of whether the assessment methods employed unfairly disadvantage one group of students relative to another. Typically, student performance is compared between males and females, and among ethnic groups. In Kentucky, only two ethnic groups are large enough to warrant statistical examination of differences in their performance: White students and Black students.

Tables 16-1 through 16-12 provide summary statistics describing the performance of groups of students by grade and content area. In addition, these tables provide statistics comparing student performance by ethnicity and gender for multiple-choice and open-ended items given in the KIRIS assessment. For each grade and content area, mean scores are presented for all students taking the test, and for groups of students considered separately by gender, by ethnicity, and by gender and ethnicity. Mean scores for each group of students are compared against the mean for all students, with the sign of Z(Mean) indicating whether a group of students scores higher (positive) or lower (negative) than all students; the magnitude of Z(Mean) indicates the extent to which students in a group score higher or lower than all students. Of principal interest in these tables is the row labeled Z(OR-MC), where positive values indicate that a group of students scores higher relative to all students on open-response items than on multiple-choice items.

RESULTS

A clear pattern of results emerges in examining these tables: females perform relatively better on open-response than on multiple-choice items. This is especially true for Black females.

TABLE 16-1
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 4 READING

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	MC	O-R	MC
Total	22031	22031	2281	2281	20619	20619	2179	2179	24312	24312	22798	22798	42650	42650	4460	4460	47110	47110
Mean	9.12	14.00	7.72	11.39	10.13	14.26	8.61	11.85	8.99	13.76	9.98	14.03	9.61	14.13	8.15	11.62	9.47	13.89
Variance	10.40	18.65	9.83	19.74	10.45	17.29	10.18	18.01	10.52	19.33	10.63	17.86	10.68	18.01	10.19	18.95	10.81	18.64
StdDev	3.23	4.32	3.13	4.44	3.23	4.16	3.19	4.24	3.24	4.40	3.26	4.23	3.27	4.24	3.19	4.35	3.29	4.32
Z(mean)	-0.11	0.03	-0.53	-0.58	0.20	0.09	-0.26	-0.47	-0.15	-0.03	0.16	0.03	0.04	0.06	-0.40	-0.53	0.00	0.00
Z(SD)	0.98	1.00	0.95	1.03	0.98	0.96	0.97	0.98	0.99	1.02	0.99	0.98	0.99	1.01	0.97	1.01	1.00	1.00
Z(OR-MC)	-0.13		0.05		0.11		0.21		-0.11		0.12		-0.02		0.13		0.00	

TABLE 16-2
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 4 SOCIAL STUDIES

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	22031	22031	2281	2281	20619	20619	2179	2179	24312	24312	22798	22798	42650	42650	4460	4460	47110	47110
Mean	8.62	12.16	7.24	9.84	9.09	12.22	7.73	10.05	8.49	11.94	8.96	12.01	8.85	12.19	7.48	9.94	8.72	11.97
Variance	11.61	14.52	10.79	13.83	11.46	13.47	10.04	12.48	11.70	14.91	11.49	13.79	11.60	14.01	10.48	13.17	11.65	14.37
StdDev	3.41	3.81	3.29	3.72	3.39	3.67	3.17	3.53	3.42	3.86	3.39	3.71	3.41	3.74	3.24	3.63	3.41	3.79
Z(mean)	-0.03	0.05	-0.43	-0.56	0.11	0.06	-0.29	-0.51	-0.07	-0.01	0.07	0.01	0.04	0.06	-0.36	-0.54	0.00	0.00
Z(SD)	1.00	1.01	0.96	0.98	0.99	0.97	0.93	0.93	1.00	1.02	0.99	0.98	1.00	0.99	0.95	0.96	1.00	1.00
Z(OR-MC)	-0.08		0.13		0.05		0.22		-0.06		0.06		-0.02		0.17		0.00	

TABLE 16-3

COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 4 MATH

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	22031	22031	2281	2281	20619	20619	2179	2179	24312	24312	22798	22798	42650	42650	4460	4460	47110	47110
Mean	8.68	12.25	6.12	9.86	8.88	12.17	6.34	9.93	8.44	12.03	8.64	11.95	8.78	12.21	6.22	9.89	8.54	11.99
Variance	20.54	16.40	16.69	14.94	20.62	15.39	16.44	15.27	20.73	16.74	20.78	15.81	20.59	15.91	16.58	15.10	20.76	16.29
StdDev	4.53	4.05	4.09	3.86	4.54	3.92	4.05	3.91	4.55	4.09	4.56	3.98	4.54	3.99	4.07	3.89	4.56	4.04
Z(mean)	0.03	0.06	-0.53	-0.53	0.08	0.04	-0.48	-0.51	-0.02	0.01	0.02	-0.01	0.05	0.05	-0.51	-0.52	0.00	0.00
Z(SD)	0.99	1.00	0.90	0.96	1.00	0.97	0.89	0.97	1.00	1.01	1.00	0.99	1.00	0.99	0.89	0.96	1.00	1.00
Z(OR-MC)	-0.03		0.00		0.03		0.03		-0.03		0.03		0.00		0.01		0.00	

TABLE 16-4

COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 4 SCIENCE

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	22031	22031	2281	2281	20619	20619	2179	2179	24312	24312	22798	22798	42650	42650	4460	4460	47110	47110
Mean	9.83	11.72	7.81	9.28	10.10	11.31	8.06	9.21	9.64	11.49	9.90	11.11	9.96	11.52	7.93	9.24	9.77	11.31
Variance	12.93	13.49	13.14	11.36	13.08	11.93	12.14	10.51	13.29	13.60	13.35	12.18	13.02	12.78	12.66	10.94	13.34	13.05
StdDev	3.60	3.67	3.62	3.37	3.62	3.45	3.48	3.24	3.65	3.71	3.65	3.49	3.61	3.57	3.56	3.31	3.65	3.61
Z(mean)	0.02	0.11	-0.54	-0.56	0.09	0.00	-0.47	-0.58	-0.03	0.05	0.04	-0.05	0.05	0.06	-0.50	-0.57	0.00	0.00
Z(SD)	0.98	1.02	0.99	0.93	0.99	0.96	0.95	0.90	1.00	1.03	1.00	0.97	0.99	0.99	0.97	0.92	1.00	1.00
Z(OR-MC)	-0.10		0.03		0.09		0.11		-0.09		0.09		-0.01		0.07		0.00	

TABLE 16-5
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 8 READING

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	23251	23251	2300	2300	22037	22037	2237	2237	25551	25551	24274	24274	45288	45288	4537	4537	49825	49825
Mean	10.26	12.78	8.35	10.52	11.73	14.07	9.83	12.14	10.09	12.58	11.55	13.89	10.97	13.41	9.08	11.32	10.80	13.22
Variance	13.46	19.44	13.41	19.37	11.35	15.81	12.56	18.18	13.75	19.85	11.77	16.34	12.98	18.09	13.53	19.43	13.32	18.57
StdDev	3.67	4.41	3.66	4.40	3.37	3.98	3.54	4.26	3.71	4.46	3.43	4.04	3.60	4.25	3.68	4.41	3.65	4.31
Z(mean)	-0.15	-0.10	-0.67	-0.63	0.25	0.20	-0.27	-0.25	-0.20	-0.15	0.21	0.16	0.05	0.04	-0.47	-0.44	0.00	0.00
Z(SD)	1.01	1.02	1.00	1.02	0.92	0.92	0.97	0.99	1.02	1.03	0.94	0.94	0.99	0.99	1.01	1.02	1.00	1.00
Z(OR-MC)	-0.05		-0.04		0.06		-0.02		-0.05		0.05		0.00		-0.03		0.00	

TABLE 16-6
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 8 MATH

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	23251	23251	2300	2300	22037	22037	2237	2237	25551	25551	24274	24274	45288	45288	4537	4537	49825	49825
Mean	7.37	10.02	4.72	7.26	7.75	9.49	5.17	7.11	7.14	9.77	7.51	9.27	7.56	9.76	4.94	7.18	7.32	9.53
Variance	20.66	18.37	15.88	11.86	19.72	16.67	15.26	10.92	20.80	18.41	19.87	16.61	20.24	17.61	15.62	11.40	20.38	17.59
StdDev	4.55	4.29	3.98	3.44	4.44	4.08	3.91	3.30	4.56	4.29	4.46	4.08	4.50	4.20	3.95	3.38	4.51	4.19
Z(mean)	0.01	0.12	-0.58	-0.54	0.10	-0.01	-0.48	-0.58	-0.04	0.06	0.04	-0.06	0.05	0.06	-0.53	-0.56	0.00	0.00
Z(SD)	1.01	1.02	0.88	0.82	0.98	0.97	0.87	0.79	1.01	1.02	0.99	0.97	1.00	1.00	0.88	0.81	1.00	1.00
Z(OR-MC)	-0.11		-0.03		0.10		0.10		-0.10		0.10		0.00		0.03		0.00	

TABLE 16-7
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY GRADE 8 SOCIAL STUDIES

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	23251	23251	2300	2300	22037	22037	2237	2237	25551	25551	24274	24274	45288	45288	4537	4537	49825	49825
Mean	7.98	11.73	6.43	9.64	9.02	11.76	7.57	9.90	7.84	11.54	8.89	11.59	8.49	11.74	6.99	9.77	8.35	11.56
Variance	12.16	17.50	11.15	14.31	11.29	14.82	11.14	13.67	12.26	17.57	11.45	15.00	12.01	16.20	11.46	14.01	12.14	16.32
StdDev	3.49	4.18	3.34	3.78	3.36	3.85	3.34	3.70	3.50	4.19	3.38	3.87	3.46	4.02	3.39	3.74	3.48	4.04
Z(mean)	-0.11	0.04	-0.55	-0.48	0.19	0.05	-0.23	-0.41	-0.15	-0.01	0.15	0.01	0.04	0.04	-0.39	-0.44	0.00	0.00
Z(SD)	1.00	1.04	0.96	0.94	0.96	0.95	0.96	0.92	1.01	1.04	0.97	0.96	0.99	1.00	0.97	0.93	1.00	1.00
Z(OR-MC)	-0.15		-0.07		0.14		0.19		-0.14		0.15		-0.01		0.05		0.00	

TABLE 16-8
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 8 SCIENCE

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	23251	23251	2300	2300	22037	22037	2237	2237	25551	25551	24274	24274	45288	45288	4537	4537	49825	49825
Mean	6.14	10.07	4.30	8.06	6.03	9.73	4.58	8.01	5.98	9.89	5.89	9.57	6.09	9.91	4.44	8.04	5.94	9.74
Variance	11.18	11.17	8.61	9.14	9.88	9.47	8.07	7.94	11.23	11.32	9.89	9.58	10.55	10.37	8.36	8.55	10.58	10.50
StdDev	3.34	3.34	2.93	3.02	3.14	3.08	2.84	2.82	3.35	3.36	3.14	3.10	3.25	3.22	2.89	2.92	3.25	3.24
Z(mean)	0.06	0.10	-0.50	-0.52	0.03	0.00	-0.42	-0.53	0.01	0.05	-0.01	-0.05	0.05	0.05	-0.46	-0.52	0.00	0.00
Z(SD)	1.03	1.03	0.90	0.93	0.97	0.95	0.87	0.87	1.03	1.04	0.97	0.96	1.00	0.99	0.89	0.90	1.00	1.00
Z(OR-MC)	-0.04		0.01		0.03		0.11		-0.04		0.04		-0.01		0.06		0.00	

TABLE 16-9
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 12 READING

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	16515	16515	1406	1406	16206	16206	1522	1522	17921	17921	17728	17728	32721	32721	2928	2928	35649	35649
Mean	8.08	12.78	6.58	10.68	9.50	13.96	7.85	11.70	7.96	12.62	9.36	13.76	8.78	13.36	7.24	11.21	8.66	13.19
Variance	9.65	15.25	7.96	15.15	9.73	12.28	8.47	14.06	9.68	15.56	9.84	12.83	10.19	14.12	8.63	14.84	10.24	14.53
StdDev	3.11	3.90	2.82	3.89	3.12	3.50	2.91	3.75	3.11	3.94	3.14	3.58	3.19	3.76	2.94	3.85	3.20	3.81
Z(mean)	-0.18	-0.11	-0.65	-0.66	0.26	0.20	-0.25	-0.39	-0.22	-0.15	0.22	0.15	0.04	0.05	-0.44	-0.52	0.00	0.00
Z(SD)	0.97	1.02	0.88	1.02	0.97	0.92	0.91	0.98	0.97	1.03	0.98	0.94	1.00	0.99	0.92	1.01	1.00	1.00
Z(OR-MC)	-0.07		0.01		0.06		0.14		-0.07		0.07		-0.01		0.08		0.00	

TABLE 16-10
COMPARISON OF MEAN SCORES ON OPEN-RESPONSE AND MULTIPLE-CHOICE TESTS BY GENDER AND ETHNICITY, GRADE 12 SOCIAL STUDIES

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	16515	16515	1406	1406	16206	16206	1522	1522	17921	17921	17728	17728	32721	32721	2928	2928	35649	35649
Mean	8.63	11.13	6.82	9.32	9.41	11.01	7.56	9.68	8.49	10.99	9.25	10.89	9.02	11.07	7.20	9.50	8.87	10.94
Variance	14.50	15.56	13.13	14.24	13.05	12.60	12.06	11.90	14.63	15.70	13.23	12.68	13.93	14.10	12.71	13.05	14.08	14.20
StdDev	3.81	3.95	3.62	3.77	3.61	3.55	3.47	3.45	3.82	3.96	3.64	3.56	3.73	3.75	3.56	3.61	3.75	3.77
Z(mean)	-0.06	0.05	-0.55	-0.43	0.15	0.02	-0.35	-0.33	-0.10	0.01	0.10	-0.01	0.04	0.03	-0.44	-0.38	0.00	0.00
Z(SD)	1.01	1.05	0.97	1.00	0.96	0.94	0.93	0.92	1.02	1.05	0.97	0.94	0.99	1.00	0.95	0.96	1.00	1.00
Z(OR-MC)	-0.11		-0.12		0.13		-0.01		-0.11		0.12		0.01		-0.06		0.00	

Table 16-11
Comparison of Mean Scores on Open-Response and Multiple-Choice Tests by Gender and Ethnicity, Grade 12 Science

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	16515	16515	1406	1406	16206	16206	1522	1522	17921	17921	17728	17728	32721	32721	2928	2928	35649	35649
Mean	7.31	10.59	4.89	8.04	7.50	9.78	5.38	7.69	7.12	10.39	7.31	9.60	7.40	10.19	5.14	7.85	7.22	10.00
Variance	13.13	16.78	10.75	12.38	11.21	13.22	9.85	10.83	13.36	16.91	11.44	13.36	12.18	15.18	10.34	11.60	12.42	15.30
StdDev	3.62	4.10	3.28	3.52	3.35	3.64	3.14	3.29	3.66	4.11	3.38	3.66	3.49	3.90	3.22	3.41	3.52	3.91
Z(mean)	0.03	0.15	-0.66	-0.50	0.08	-0.06	-0.52	-0.59	-0.03	0.10	0.03	-0.10	0.05	0.05	-0.59	-0.55	0.00	0.00
Z(SD)	1.03	1.05	0.93	0.90	0.95	0.93	0.89	0.84	1.04	1.05	0.96	0.93	0.99	1.00	0.91	0.87	1.00	1.00
Z(OR-MC)	-0.13		-0.16		0.13		0.07		-0.13		0.13		0.00		-0.04		0.00	

Table 16-12
Comparison of Mean Scores on Open-Response and Multiple-Choice Tests by Gender and Ethnicity, Grade 12 Math

	White Male		Black Male		White Female		Black Female		Male		Female		White		Black		Total	
Score	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C	O-R	M-C
Total	16515	16515	1406	1406	16206	16206	1522	1522	17921	17921	17728	17728	32721	32721	2928	2928	35649	35649
Mean	5.72	9.39	3.75	7.17	5.82	8.92	3.94	7.16	5.57	9.21	5.66	8.77	5.77	9.16	3.85	7.16	5.61	8.99
Variance	12.36	15.52	7.45	10.24	10.02	13.17	7.50	11.35	12.26	15.46	10.08	13.26	11.21	14.41	7.48	10.81	11.18	14.41
StdDev	3.52	3.94	2.73	3.20	3.17	3.63	2.74	3.37	3.50	3.93	3.18	3.64	3.35	3.80	2.74	3.29	3.34	3.80
Z(mean)	0.03	0.10	-0.56	-0.48	0.06	-0.02	-0.50	-0.48	-0.01	0.06	0.01	-0.06	0.05	0.04	-0.53	-0.48	0.00	0.00
Z(SD)	1.05	1.04	0.82	0.84	0.95	0.96	0.82	0.89	1.05	1.04	0.95	0.96	1.00	1.00	0.82	0.87	1.00	1.00
Z(OR-MC)	-0.07		-0.08		0.08		-0.02		-0.07		0.07		0.00		-0.04		0.00	

Appendix 7-1
Principal's and District Assessment Coordinator's Certification of Proper Test Administration Forms

KENTUCKY INSTRUCTIONAL RESULTS INFORMATION SYSTEM (KIRIS)
TRANSITIONAL ASSESSMENT
1993-94

PRINCIPAL'S CERTIFICATION OF PROPER TEST ADMINISTRATION
Grade 4

District Name and Code:

School Name and Code:

1. Number of students enrolled in the fourth grade as of April 18, 1994 ______ (1.)
    2. Number of students tested ______ (2.)
    3. Number of students excluded (See note below.) ______ (3.)
    4. Number of students not tested (See note below.) ______ (4.)
5. Total number of used student response booklets ______ (5.)
   (Lines 2. + 3. + 4. = 5.) If Line 5. does not equal Line 1., please explain why at the bottom of this page.
6. Number of unused student response booklets ______ (6.)
7. Total number of student response booklets being returned to Advanced Systems ______ (7.)
   (Lines 5. + 6. = 7.)
8. Total number of student response booklets sent to you ______ (8.)
   If Line 7. does not equal Line 8., please explain the discrepancy on the bottom of this page.

NOTE: **You must complete test booklets for all students whether they are tested or not.** Student response booklets for those students who are excluded from testing for special education,* limited-English speaking, or medical reasons in accordance with Department of Education policy should have the appropriate reason for exclusion indicated in the proper section on page 1.

* For this purpose, special education students are those eligible for and participating in the alternate portfolio assessment.

The information provided above is correct to the best of my knowledge. I have also read and initialed the reverse side of this form.

______________________________ ______________
Principal's Signature Date

Discrepancy Notes:

I certify that:

- the information recorded on page 1 of all student response booklets and this form is correct to the best of my knowledge;
- every effort was made to administer the full assessment to all eligible fourth-grade students;
- the testing was done in accordance with the directions in the *Instruction Manual for Test Administrators* and the *Instruction Manual for Assessment Coordinators*;
- I have checked to see that a student response booklet is assigned to every student enrolled on April 18, 1994, whether he/she was tested or not, and all appropriate sections on page 1 of his/her student response booklet have been properly completed;
- I understand that the KIRIS student assessment test booklets are secure materials and, to the best of my knowledge, have not been reproduced by anyone in my school;
- test booklets were distributed randomly as directed in the administrator's manual;
- to the best of my knowledge, no personnel or student in the school shared answers to test questions or otherwise gave out information that would lead to inappropriate test administration;
- test administrators have read and signed a copy of *Appropriate Assessment Practices* and the copies are maintained on file;
- any authorized photocopying of **student responses to the first four common open-response items** has been done in accordance with procedures provided by the District Assessment Coordinator;
- **all** test-related materials are being returned to Advanced Systems as directed;
- assessment materials revealing assessment content that inadvertently remain in the school after such material was to have been returned to Advanced Systems will be secured and returned immediately upon being found; and
- I shall inform the Department and the local district assessment coordinator immediately if any inappropriate uses of assessment materials are discovered within this school.

____

Principal's Initials

KENTUCKY INSTRUCTIONAL RESULTS INFORMATION SYSTEM (KIRIS)
TRANSITIONAL ASSESSMENT
1993-94

DISTRICT ASSESSMENT COORDINATOR'S CERTIFICATION OF PROPER TEST ADMINISTRATION
Grade 12

District Name and Code:

I certify that:

- the information recorded on page 1 of all student response booklets and this form is correct to the best of my knowledge;
- every effort was made to administer the full assessment to all eligible twelfth-grade students;
- the testing was done in accordance with the directions in the *Instruction Manual for Test Administrators* and the *Instruction Manual for Assessment Coordinators*;
- I have checked to see that a student response booklet is assigned to every student enrolled on February 7, 1994, whether he/she was tested or not, and all appropriate sections on page 1 of his/her student response booklet have been properly completed;
- I understand that the KIRIS student assessment test booklets are secure materials and, to the best of my knowledge, have not been reproduced by anyone in my district;
- test booklets were distributed randomly as directed in the administrator's manual;
- to the best of my knowledge, no personnel or student in the district shared answers to test questions or otherwise gave out information that would lead to inappropriate test administration;
- test administrators have read and signed a copy of *Appropriate Assessment Practices* and the copies are maintained on file;
- any authorized photocopying of **student responses to the first four common open-ended items** has been done in accordance with procedures provided by the Kentucky Department of Education;
- all test-related materials are being returned to Advanced Systems as directed;
- assessment materials revealing assessment content that inadvertently remain in the school after such material was to have been returned to Advanced Systems will be secured and returned immediately upon being found; and
- I shall inform the Department immediately if any inappropriate uses of assessment materials are discovered within this district.

District Assessment Coordinator's Signature

Date

Appendix 7-2
Student Answer Document

KIRIS 1993–1994 WRITING PORTFOLIO SCORE FORM

GRADE 4

0400347

REMOVE LABEL AND PLACE ON STUDENT PORTFOLIO

PLEASE DO NOT WRITE IN THIS AREA

● ● ○ ● ● ○ ● ● ● ● ○ ● ● ○ ○ ○ ○ ● ● ○ ○ ○ ○ ○

0400347

Mark Reflex® by NCS EM-154300:321 Printed in U.S.A.

MARKING INSTRUCTIONS

CORRECT MARK

●

INCORRECT MARKS

⊘ ⊗ ◒ ⊙

- Use No. 2 pencil only.
- Do not use ink, ballpoint, or felt tip pens.
- Make solid marks that fill the circle completely.
- Erase cleanly any marks you wish to change.
- Make no stray marks on this form.
- Do not fold, tear, or mutilate this form.

DISTRICT NAME: ______________________

SCHOOL NAME: ______________________

STUDENT NAME: ______________________

STUDENT DATE OF BIRTH: ______________________
(MONTH) (DAY) (YEAR)

HOLISTIC SCORE

DARKEN 1 BUBBLE

○ Distinguished

○ Proficient

○ Apprentice

○ Novice

○ Incomplete

○ Blank

LITHOCODE OVERRIDE

⓪	⓪	⓪	⓪	⓪	⓪	⓪
①	①	①	①	①	①	①
②	②	②	②	②	②	②
③	③	③	③	③	③	③
④	④	④	④	④	④	④
⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨	⑨	⑨	⑨

STUDENT INFORMATION

Has this student been enrolled in a Kentucky Public School for fewer than 100 days?

○ Yes ○ No

If you will not be reporting a score for this student, complete where appropriate:

○ Eligible for alternative portfolio per Department of Education policy

○ Foreign Exchange Student

○ Non-English speaking exclusion per Department of Education policy

○ Moved

○ Other, explain: ______________________

ANALYTIC ANNOTATION

DARKEN 2 BUBBLES IN ANY COMBINATION

COMMENDATIONS		NEEDS	
P/A-X	○	P/A-J	○
P/A-Y	○	P/A-K	○
I/S-X	○	I/S-J	○
I/S-Y	○	I/S-K	○
OX	○	OJ	○
OY	○	OK	○
SX	○	SJ	○
SY	○	SK	○
WX	○	WJ	○
WY	○	WK	○
SF-X	○	SF-J	○
SF-Y	○	SF-K	○

KIRIS 1993–1994 MATHEMATICS PORTFOLIO SCORE FORM

KENTUCKY DEPARTMENT OF EDUCATION

0400347

REMOVE LABEL AND PLACE ON STUDENT PORTFOLIO

GRADE 4

PLEASE DO NOT WRITE IN THIS AREA

MARKING INSTRUCTIONS

CORRECT MARK

●

INCORRECT MARKS

- Use No. 2 pencil only.
- Do not use ink, ballpoint, or felt tip pens.
- Make solid marks that fill the circle completely.
- Erase cleanly any marks you wish to change.
- Make no stray marks on this form.
- Do not fold, tear, or mutilate this form.

DISTRICT NAME: ______________________________

SCHOOL NAME: ______________________________

STUDENT NAME: ______________________________

STUDENT DATE OF BIRTH: ______________________
(MONTH) (DAY) (YEAR)

HOLISTIC SCORE

DARKEN 1 BUBBLE

○ Distinguished

○ Proficient

○ Apprentice

○ Novice

○ Incomplete

○ Blank

LITHOCODE OVERRIDE

⓪	⓪	⓪	⓪	⓪	⓪	⓪
①	①	①	①	①	①	①
②	②	②	②	②	②	②
③	③	③	③	③	③	③
④	④	④	④	④	④	④
⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨	⑨	⑨	⑨

Mark Reflex® by NCS EM-155526:321 Printed in U.S.A.

0400347

273

STUDENT INFORMATION

Has this student been enrolled in a Kentucky Public School for fewer than 100 days?

○ Yes ○ No

If you will not be reporting a score for this student, complete where appropriate:

○ Eligible for alternative portfolio per Department of Education policy

○ Foreign Exchange Student

○ Non-English speaking exclusion per Department of Education policy

○ Moved

○ Other, explain: ______________________

BREADTH OF ENTRIES

TYPES

○ Investigations/Disclosures
○ Applications
○ Non-routine Problems
○ Projects
○ Interdisciplinary
○ Writing

TOOLS

○ Calculators
○ Computer and Technology
○ Models/Manipulatives
○ Measurement Instruments
○ Others

CORE CONCEPTS AND PRINCIPLES

○ Number
○ Mathematical Procedures
○ Space and Dimensionality
○ Measurement
○ Change
○ Mathematical Structure
○ Data

274

STUDENT RESPONSE BOOKLET

Grade 4

DISTRICT NAME: ______________________________

SCHOOL NAME: ______________________________

STUDENT NAME: ______________________________

SEX

○ M

○ F

ETHNICITY

○ White (Non-Hispanic)

○ Black (Non-Hispanic)

○ Hispanic

○ Asian

○ Other

BIRTH DATE

Mo.	Day		Yr.	
JAN ○				
FEB ○				
MAR ○	⓪	⓪		⓪
APR ○	①	①		①
MAY ○	②	②		②
JUNE ○	③	③		③
JULY ○		④		④
AUG ○		⑤		⑤
SEPT ○		⑥		⑥
OCT ○		⑦		⑦
NOV ○		⑧	⑧	⑧
DEC ○		⑨	⑨	⑨

LAST NAME														FIRST NAME											MI
○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ
Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ
Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ
Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ
Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ	Ⓕ
Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ	Ⓖ
Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ	Ⓗ
Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ	Ⓘ
Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ	Ⓙ
Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ	Ⓚ
Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ	Ⓛ
Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ	Ⓜ
Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ	Ⓝ
Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ	Ⓞ
Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ	Ⓟ
Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ	Ⓠ
Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ	Ⓡ
Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ
Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ	Ⓣ
Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ	Ⓤ
Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ	Ⓥ
Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ	Ⓦ
Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ	Ⓧ
Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ
Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ	Ⓩ

Mark Reflex® by NCS EM 159528 321 Printed in U.S.A.

PAGE 1

TO BE COMPLETED BY SCHOOL PERSONNEL

1 IF THIS STUDENT WAS NOT TESTED, COMPLETE WHERE APPROPRIATE:

○ Eligible for alternate portfolio in accordance with Department of Education policy

○ Non-English speaking exclusion in accordance with Department of Education policy

○ Medical exclusion (signed doctor's note attached)

○ Foreign Exchange Student

○ Absent

○ Moved

○ Other, please explain: ______________________________

____

3 Does this student have an identified educational handicap? If yes, mark the primary handicapping condition.

○ Autism

○ Deaf/Blind

○ Multiple Disabilities

○ Emotional/Behavioral Disabilities

○ Mild Mental Disabilities

○ Physical Disabilities/ Orthopedically Impaired

○ Other Health Impaired Disabilities

○ Traumatic Brain Injury

○ Hearing Impaired

○ Visual Disabilities

○ Communication Disabilities/Speech-Language Disabilities

○ Functional Mental Disabilities

○ Specific Learning Disabilities

2 Does this student participate in Chapter 1?

○ Yes ○ No

If yes, please indicate the program that is serving this student. Mark only one.

○ Chapter 1 Basic Program

○ Migrant Program

○ Both Chapter 1 Basic and Migrant Programs

For the Chapter 1 Basic Program, mark only one.

○ Reading skills only

○ Mathematics skills only

○ Reading and Mathematics skills

For the Migrant Program, mark only one.

○ Served and classified currently

○ Served and classified formerly first year

○ Served and classified formerly beyond first year

4 Has this student participated in Extended School Services more than 5 days during this school year?

○ Yes ○ No

STUDENT QUESTIONNAIRE

1	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	15	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
2	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	16	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
3	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	17	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
4	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	18	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
5	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	19	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
6	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	20	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
7	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	21	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
8	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	22	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
9	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	23	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
10	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	24	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
11	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	25	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
12	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	26	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
13	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	27	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
14	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	28	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ

STUDENT PERMISSION

May we consider using your responses in teacher publications? Your name will not be used. If we may use your responses, please bubble yes.

○ YES ○ NO

OVERRIDE

0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9

277

PLEASE DO NOT WRITE IN THIS AREA

●●○●●○●●●●○●●○○○○●●○○○○○ 0400347

READING
SESSION A

GRADE 4

FORM
① ② ③ ④
⑤ ⑥ ⑦ ⑧
⑨ ⑩ ⑪ ⑫

⓪	⓪	⓪
①	①	①
②	②	②
③	③	③
④	④	④
⑤	⑤	⑤
⑥	⑥	⑥
⑦	⑦	⑦
⑧	⑧	⑧
⑨	⑨	⑨

NAME: ______________________________

1 Ⓐ Ⓑ Ⓒ Ⓓ
2 Ⓐ Ⓑ Ⓒ Ⓓ
3 Ⓐ Ⓑ Ⓒ Ⓓ
4 Ⓐ Ⓑ Ⓒ Ⓓ
5 Ⓐ Ⓑ Ⓒ Ⓓ
6 Ⓐ Ⓑ Ⓒ Ⓓ

7

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

8 Ⓐ Ⓑ Ⓒ Ⓓ
9 Ⓐ Ⓑ Ⓒ Ⓓ
10 Ⓐ Ⓑ Ⓒ Ⓓ
11 Ⓐ Ⓑ Ⓒ Ⓓ

12

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

278

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

Mark Reflex® by NCS EM-156529:321 Printed in U.S.A.

0400347

READING (continued)

13 Ⓐ Ⓑ Ⓒ Ⓓ	15 Ⓐ Ⓑ Ⓒ Ⓓ	17 Ⓐ Ⓑ Ⓒ Ⓓ
14 Ⓐ Ⓑ Ⓒ Ⓓ	16 Ⓐ Ⓑ Ⓒ Ⓓ	18 Ⓐ Ⓑ Ⓒ Ⓓ

19

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

SESSION B

20 Ⓐ Ⓑ Ⓒ Ⓓ	22 Ⓐ Ⓑ Ⓒ Ⓓ
21 Ⓐ Ⓑ Ⓒ Ⓓ	23 Ⓐ Ⓑ Ⓒ Ⓓ

24

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

READING (continued)

25 Ⓐ Ⓑ Ⓒ Ⓓ
26 Ⓐ Ⓑ Ⓒ Ⓓ
27 Ⓐ Ⓑ Ⓒ Ⓓ
28 Ⓐ Ⓑ Ⓒ Ⓓ
29 Ⓐ Ⓑ Ⓒ Ⓓ

30

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

31 Ⓐ Ⓑ Ⓒ Ⓓ
32 Ⓐ Ⓑ Ⓒ Ⓓ
33 Ⓐ Ⓑ Ⓒ Ⓓ

34

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

280

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

35 Ⓐ Ⓑ Ⓒ Ⓓ
36 Ⓐ Ⓑ Ⓒ Ⓓ
37 Ⓐ Ⓑ Ⓒ Ⓓ
38 Ⓐ Ⓑ Ⓒ Ⓓ
39 Ⓐ Ⓑ Ⓒ Ⓓ

40

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

EXTRA SPACE FOR OPEN-RESPONSE ____

MATHEMATICS

SESSION A

GRADE 4

FORM
① ② ③ ④
⑤ ⑥ ⑦ ⑧
⑨ ⑩ ⑪ ⑫

⓪ ⓪ ⓪
① ① ①
② ② ②
③ ③ ③
④ ④ ④
⑤ ⑤ ⑤
⑥ ⑥ ⑥
⑦ ⑦ ⑦
⑧ ⑧ ⑧
⑨ ⑨ ⑨

NAME: ____________________

1 Ⓐ Ⓑ Ⓒ Ⓓ	5 Ⓐ Ⓑ Ⓒ Ⓓ	9 Ⓐ Ⓑ Ⓒ Ⓓ	13 Ⓐ Ⓑ Ⓒ Ⓓ
2 Ⓐ Ⓑ Ⓒ Ⓓ	6 Ⓐ Ⓑ Ⓒ Ⓓ	10 Ⓐ Ⓑ Ⓒ Ⓓ	14 Ⓐ Ⓑ Ⓒ Ⓓ
3 Ⓐ Ⓑ Ⓒ Ⓓ	7 Ⓐ Ⓑ Ⓒ Ⓓ	11 Ⓐ Ⓑ Ⓒ Ⓓ	15 Ⓐ Ⓑ Ⓒ Ⓓ
4 Ⓐ Ⓑ Ⓒ Ⓓ	8 Ⓐ Ⓑ Ⓒ Ⓓ	12 Ⓐ Ⓑ Ⓒ Ⓓ	16 Ⓐ Ⓑ Ⓒ Ⓓ

17

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

18

282

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

19

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

SESSION B

20 Ⓐ Ⓑ Ⓒ Ⓓ
21 Ⓐ Ⓑ Ⓒ Ⓓ
22 Ⓐ Ⓑ Ⓒ Ⓓ
23 Ⓐ Ⓑ Ⓒ Ⓓ
24 Ⓐ Ⓑ Ⓒ Ⓓ
25 Ⓐ Ⓑ Ⓒ Ⓓ
26 Ⓐ Ⓑ Ⓒ Ⓓ
27 Ⓐ Ⓑ Ⓒ Ⓓ
28 Ⓐ Ⓑ Ⓒ Ⓓ
29 Ⓐ Ⓑ Ⓒ Ⓓ
30 Ⓐ Ⓑ Ⓒ Ⓓ
31 Ⓐ Ⓑ Ⓒ Ⓓ

32

283

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

33

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

34

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

35

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

EXTRA SPACE FOR OPEN-RESPONSE ____

SCIENCE

SESSION A

GRADE 4

⓪	⓪	⓪
①	①	①
②	②	②
③	③	③
④	④	④
⑤	⑤	⑤
⑥	⑥	⑥
⑦	⑦	⑦
⑧	⑧	⑧
⑨	⑨	⑨

FORM
① ② ③ ④
⑤ ⑥ ⑦ ⑧
⑨ ⑩ ⑪ ⑫

NAME: ____________________

1 Ⓐ Ⓑ Ⓒ Ⓓ	5 Ⓐ Ⓑ Ⓒ Ⓓ	9 Ⓐ Ⓑ Ⓒ Ⓓ	13 Ⓐ Ⓑ Ⓒ Ⓓ
2 Ⓐ Ⓑ Ⓒ Ⓓ	6 Ⓐ Ⓑ Ⓒ Ⓓ	10 Ⓐ Ⓑ Ⓒ Ⓓ	14 Ⓐ Ⓑ Ⓒ Ⓓ
3 Ⓐ Ⓑ Ⓒ Ⓓ	7 Ⓐ Ⓑ Ⓒ Ⓓ	11 Ⓐ Ⓑ Ⓒ Ⓓ	15 Ⓐ Ⓑ Ⓒ Ⓓ
4 Ⓐ Ⓑ Ⓒ Ⓓ	8 Ⓐ Ⓑ Ⓒ Ⓓ	12 Ⓐ Ⓑ Ⓒ Ⓓ	16 Ⓐ Ⓑ Ⓒ Ⓓ

17

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

18

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

256

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

19

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

SESSION B

20 Ⓐ Ⓑ Ⓒ Ⓓ
21 Ⓐ Ⓑ Ⓒ Ⓓ
22 Ⓐ Ⓑ Ⓒ Ⓓ
23 Ⓐ Ⓑ Ⓒ Ⓓ
24 Ⓐ Ⓑ Ⓒ Ⓓ
25 Ⓐ Ⓑ Ⓒ Ⓓ
26 Ⓐ Ⓑ Ⓒ Ⓓ
27 Ⓐ Ⓑ Ⓒ Ⓓ
28 Ⓐ Ⓑ Ⓒ Ⓓ
29 Ⓐ Ⓑ Ⓒ Ⓓ
30 Ⓐ Ⓑ Ⓒ Ⓓ
31 Ⓐ Ⓑ Ⓒ Ⓓ

32

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

33

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

34

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

35

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

EXTRA SPACE FOR OPEN-RESPONSE ____

SOCIAL STUDIES

SESSION A

GRADE 4

⓪	⓪	⓪
①	①	①
②	②	②
③	③	③
④	④	④
⑤	⑤	⑤
⑥	⑥	⑥
⑦	⑦	⑦
⑧	⑧	⑧
⑨	⑨	⑨

FORM
① ② ③ ④
⑤ ⑥ ⑦ ⑧
⑨ ⑩ ⑪ ⑫

NAME: ______________________________

1	Ⓐ Ⓑ Ⓒ Ⓓ	5	Ⓐ Ⓑ Ⓒ Ⓓ	9	Ⓐ Ⓑ Ⓒ Ⓓ	13	Ⓐ Ⓑ Ⓒ Ⓓ
2	Ⓐ Ⓑ Ⓒ Ⓓ	6	Ⓐ Ⓑ Ⓒ Ⓓ	10	Ⓐ Ⓑ Ⓒ Ⓓ	14	Ⓐ Ⓑ Ⓒ Ⓓ
3	Ⓐ Ⓑ Ⓒ Ⓓ	7	Ⓐ Ⓑ Ⓒ Ⓓ	11	Ⓐ Ⓑ Ⓒ Ⓓ	15	Ⓐ Ⓑ Ⓒ Ⓓ
4	Ⓐ Ⓑ Ⓒ Ⓓ	8	Ⓐ Ⓑ Ⓒ Ⓓ	12	Ⓐ Ⓑ Ⓒ Ⓓ	16	Ⓐ Ⓑ Ⓒ Ⓓ

17

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

18

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

19

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

SESSION B

20 Ⓐ Ⓑ Ⓒ Ⓓ	23 Ⓐ Ⓑ Ⓒ Ⓓ	26 Ⓐ Ⓑ Ⓒ Ⓓ	29 Ⓐ Ⓑ Ⓒ Ⓓ
21 Ⓐ Ⓑ Ⓒ Ⓓ	24 Ⓐ Ⓑ Ⓒ Ⓓ	27 Ⓐ Ⓑ Ⓒ Ⓓ	30 Ⓐ Ⓑ Ⓒ Ⓓ
22 Ⓐ Ⓑ Ⓒ Ⓓ	25 Ⓐ Ⓑ Ⓒ Ⓓ	28 Ⓐ Ⓑ Ⓒ Ⓓ	31 Ⓐ Ⓑ Ⓒ Ⓓ

32

○⓪①②③④
○⓪①②③④

291

○⓪①②③④
○⓪①②③④

33

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

34

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

○ ⓪ ① ② ③ ④
○ ⓪ ① ② ③ ④

292

35

○⓪①②③④
○⓪①②③④

○⓪①②③④
○⓪①②③④

EXTRA SPACE FOR OPEN-RESPONSE _____

GRADE 4

WRITING TASK NUMBER

① ② ③ ④ ⑤ ⑥
⑦ ⑧ ⑨ ⑩ ⑪ ⑫
⑬ ⑭ ⑮ ⑯ ⑰ ⑱
⑲ ⑳ ㉑ ㉒ ㉓ ㉔

WRITING TASK BOOKLET

NAME: ______________________________

294

DO NOT WRITE IN THIS AREA

1

NE ○
OT ○
NR ○
ARB ○

HOLISTIC SCORE	
Ⓝ Ⓐ Ⓟ Ⓓ	
ANNOTATIONS	
C	**N**
P/A Ⓧ Ⓨ	P/A Ⓙ Ⓚ
I/S Ⓧ Ⓨ	I/S Ⓙ Ⓚ
O Ⓧ Ⓨ	O Ⓙ Ⓚ
S Ⓧ Ⓨ	S Ⓙ Ⓚ
W Ⓧ Ⓨ	W Ⓙ Ⓚ
SF Ⓧ Ⓨ	SF Ⓙ Ⓚ

SCORER #1

DO NOT WRITE IN THIS AREA

2

NE ○
OT ○
NR ○
ARB ○

HOLISTIC SCORE	
Ⓝ Ⓐ Ⓟ Ⓓ	
ANNOTATIONS	
C	**N**
P/A Ⓧ Ⓨ	P/A Ⓙ Ⓚ
I/S Ⓧ Ⓨ	I/S Ⓙ Ⓚ
O Ⓧ Ⓨ	O Ⓙ Ⓚ
S Ⓧ Ⓨ	S Ⓙ Ⓚ
W Ⓧ Ⓨ	W Ⓙ Ⓚ
SF Ⓧ Ⓨ	SF Ⓙ Ⓚ

SCORER #2

PLEASE DO NOT WRITE IN THIS AREA

●●○●●○●●●●○●●○○○○●●○○○○○ 0400347

KENTUCKY DEPARTMENT OF EDUCATION
To ensure for each child an internationally superior education and a love of learning

Appendix 7-3

Summary of Performance Events

Summary of 1993-94 Performance Events

Grade 4

S1 - "Swingers" Students experiment with a pendulum to determine why it swings at different speeds at different times.

S2 - "Some of its parts" Students examine parts of several machines to determine their function.

S/SS1 - "People interacting with their environment" Given different environmental conditions, students are asked to determine what living conditions would be like for people & animals.

M/S1 - "Rising waters" Students examine a series of floating balls to determine why each ball makes the water level rise a specific amount.

VS1 - "Blue ribbon worker award" Students list qualities of good workers in several different jobs and decide what makes a worker and a student "good".

AH1 - "Dance movement" Students use concepts of space, energy, and time to understand the elements of dance movements.

AH2 - "Building" By examining a series of pictures, students explore the use of geometric figures in buildings and how these shapes are reproduced in nature.

SS1 - "Forming a playground committee" After examining information about 15 people, students decide which of the people have the best qualifications to be on a playground committee.

SS2 - "Town planning" Students design an ideal town, listing the components they would need and why they chose them.

PL1 - "Dealing with stress" Students discuss different ways of dealing with stress, their pros and cons, and the methods they use to combat stress.

PL2 - "What's wrong with this picture?" After examining a picture of a kitchen, students discuss the dangers in the kitchen and ways to make it safer.

M1 - "Friendship bracelets" Students use different color beads to make as many different combinations of friendship bracelets as possible.

M/SS1 - "Planning a bicycle race course" Students must design a bicycle race course through town, making sure to include several key components.

Grade 8

AH1 - "Architecture" Students use perspective drawings and models to explore how shapes can be represented in two and three dimensions.

AH2 - "Dance forms" After viewing 2 dance performances, students discuss how the elements of dance combine to convey a message.

AH3 - "Music" Students examine how musical cues other than melody affect a piece of music.

M1 - "Geoboards" Students use a geoboard (a wooden board with raised pegs that form a grid) to create shapes with specific dimensions.

M/SS1 - "Native American removal" Students examine the forced relocation of Native American tribes, and the effect this relocation had on their survival.

S/SS1 - "River planning" Students are asked to consider the characteristics of a dam, and assess the impact of the dam they design on the surrounding terrain.

SS1 - "Fall of Rome" After learning about the fall of the Roman Empire, students compare conditions in Rome to conditions in the present-day United States.

SS2 - "Promote a Kentucky region" Students create a state motto and license plate to promote a selected area of Kentucky.

S2 - "Powder" Students are given several unidentified powdered substances. They are asked to examine and suggest a use for each .

VS1 - "Career day" Students are given data on demographics and job trends and must decide which speakers to invite to speak at a career day.

VS2 - "Teens who care" Students select from among several public service jobs and explain why they are qualified to hold one of the jobs.

PL1 - "Dear Lynn" Students explore some causes of teen peer pressure and write to a fellow student who is being pressured.

PL2 - "Safety first" Students are given the job of baby-sitters and must discuss common dangers that can befall young children.

Grade 12

SS/M1 - "Population center" Student study population information and predict future population growth and shift.

AH1 - "Judging a music contest" Students generate criteria for judging a musical competition and determine which performance was best based on those criteria

AH2 - "Artistic motion" Students examine several different works of art to determine how motion is conveyed in two and three dimensions.

AH3 - "Dramatic performances" After viewing a dramatic performance, students determine how elements combine to achieve a desired effect.

S1 - "That's the way the ball bounces" Students are given 4 different balls and are asked to compare the bouncing ability of each material.

S/SS1 - "Old problems, new solutions" Students convert an abandoned coal mine into useful land, analyzing the effect of both the mine and the land reclamation on the surrounding area.

SS1 - "State future" After analyzing reasons for Kentucky's population decline, students design an informational item to encourage people to live in Kentucky.

SS2 - "The clash of the cultures" Students contrast the lifestyles of the early settlers and the Native American tribes they encountered, and examine how differences in lifestyle led to conflict between the two.

M1 - "Population survey" Students are given survey data and asked to analyze both the results and the accuracy of the survey itself.

M2 - "Tiling" Students use tiles of various shapes to fill a grid, and extrapolate from their findings to grids of other sizes.

VS1 - "Conflict management" Students read a conflict scenario, discuss rules of conflict management, and rewrite the scenario so that no rules are broken.

VS2 - "Post-secondary job search" Given information on a high school graduate, students are asked which of several jobs most closely matches the student's goals, as well as how the student can prepare for the job.

M/S1 - "Crater simulation" Students determine the relationship between the velocity of a meteorite and the impact it makes by using the Law of Conservation of Energy.

PL1 - "Checkbook" Students are given a fictitious bank account statement and must reconcile the debits and credits.

Appendix 9-1

Locations of Open-Response Linking Items by Year

LINKING ITEMS
GRADE 4
READING

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 9	Item #7		
Common	Item #2	Form 8	Item #6		
Common	Item #3	Form 12	Item #7	Form 10	Item #4
Common	Item #4	Form 4	Item #7		
Common	Item #5	Form 6	Item #7		
Form 2	Item #7	Form 11	Item #6		
Form 4	Item #7	Form 10	Item #7		
Form 5	Item #6	Form 3	Item #7	Form 12	Item #4
Form 6	Item #6	Form 12	Item #6		
Form 6	Item #7	Form 2	Item #6	Form 8	Item #4
Form 8	Item #6	Form 7	Item #7	Form 7	Item #4
Form 11	Item #6	Form 7	Item #6		
Form 12	Item #6	Form 9	Item #6		
Form 12	Item #7	Form 3	Item #6	Form 4	Item #4
		Common	Item #1	Form 1	Item #4
		Common	Item #2	Form 5	Item #4
		Common	Item #3	Form 11	Item #4
		Common	Item #4	Form 9	Item #4
		Common	Item #5	Form 6	Item #4
		Form 1	Item #6	Form 2	Item #4

LINKING ITEMS
GRADE 4
MATHEMATICS

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 6	Item #6	Form 4	Item #1
Common	Item #2	Form 4	Item #6	Form 5	Item #1
Common	Item #3	Form 3	Item #6		
Common	Item #4	Form 7	Item #6		
Common	Item #5	Form 5	Item #6		
Form 2	Item #6	Form 2	Item #6		
Form 2	Item #7	Form 2	Item #7		
Form 4	Item #7	Form 4	Item #7		
Form 5	Item #6	Form 1	Item #6		
Form 5	Item #7	Form 5	Item #7		
Form 7	Item #7	Form 12	Item #7		
Form 8	Item #7	Form 6	Item #7		
Form 9	Item #7	Form 9	Item #7	Form 1	Item #1
Form 11	Item #6	Form 11	Item #6		
Form 12	Item #7	Form 1	Item #7		
		Common	Item #1	Form 6	Item #1
		Common	Item #2	Form 3	Item #1
		Common	Item #3	Form 9	Item #1
		Common	Item #4	Form 12	Item #1
		Common	Item #5	Form 8	Item #1
		Form 7	Item #7	Form 2	Item #1
		Form 11	Item #7	Form 10	Item #1

LINKING ITEMS
GRADE 4
SCIENCE

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 6	Item #7		
Common	Item #2	Form 7	Item #7		
Common	Item #3	Form 8	Item #6	Form 7	Item #1
Common	Item #4	Form 9	Item #7		
Common	Item #5	Form 1	Item #6		
Form 4	Item #6	Form 9	Item #6		
Form 5	Item #7	Form 5	Item #7		
Form 6	Item #7	Form 8	Item #7		
		Common	Item #1	Form 11	Item #1
		Common	Item #2	Form 5	Item #1
		Common	Item #3	Form 10	Item #1
		Common	Item #4	Form 2	Item #1
		Common	Item #5	Form 12	Item #1
		Form 1	Item #7	Form 3	Item #1
		Form 2	Item #6	Form 4	Item #1
		Form 3	Item #7	Form 6	Item #1
		Form 6	Item #6	Form 1	Item #1

LINKING ITEMS GRADE 4 SOCIAL STUDIES					
1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 3	Item #6		
Common	Item #2	Form 2	Item #6	Form 5	Item #1
Common	Item #3	Form 2	Item #7		
Common	Item #5	Form 1	Item #7		
Form 1	Item #6	Form 1	Item #6	Form 4	Item #1
Form 4	Item #6	Form 4	Item #6		
Form 5	Item #7	Form 8	Item #7		
Form 6	Item #6	Form 6	Item #6	Form 3	Item #1
Form 7	Item #6	Form 7	Item #6	Form 1	Item #1
Form 9	Item #6	Form 9	Item #6		
Form 9	Item #7	Form 3	Item #7	Form 2	Item #1
Form 10	Item #6	Form 12	Item #6		
Form 10	Item #7	Form 10	Item #7		
Form 11	Item #6	Form 11	Item #6		
Form 12	Item #7	Form 12	Item #7		
		Common	Item #1	Form 11	Item #1
		Common	Item #2	Form 9	Item #1
		Common	Item #3	Form 12	Item #1
		Common	Item #4	Form #10	Item #1
		Common	Item #5	Form #8	Item #1
		Form 4	Item #7	Form 7	Item #1

LINKING ITEMS
GRADE 8
READING

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 7	Item #7		
Common	Item #2	Form 5	Item #6	Form 5	Item #4
Common	Item #3	Form 7	Item #6	Form 7	Item #1
Common	Item #4	Form 9	Item #6		
Common	Item #5	Form 1	Item #7		
Form 1	Item #6	Form 1	Item #6		
Form 2	Item #6	Form 2	Item #6	Form 2	Item #4
Form 2	Item #7	Form 2	Item #7		
Form 4	Item #6	Form 12	Item #7		
Form 5	Item #6	Form 10	Item #7		
Form 5	Item #7	Form 6	Item #7	Form 6	Item #4
Form 8	Item #6	Form 8	Item #6		
Form 9	Item #7	Form 9	Item #7		
Form 10	Item #6	Form 6	Item #6		
Form 10	Item #7	Form 10	Item #7		
Form 11	Item #6	Form 11	Item #6		
Form 11	Item #7	Form 3	Item #7		
Form 12	Item #7	Form 4	Item #7	Form 4	Item #4
		Common	Item #1	Form 11	Item #4
		Common	Item #2	Form 1	Item #4
		Common	Item #3	Form 3	Item #4
		Common	Item #4	Form 12	Item #4
		Common	Item #5	Form 9	Item #4
		Form 10	Item #6	Form 10	Item #4

LINKING ITEMS
GRADE 8
MATHEMATICS

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 4	Item #7		
Common	Item #2	Form 11	Item #7		
Common	Item #3	Form 4	Item #6	Form 4	Item #1
Common	Item #4	Form 3	Item #6	Form 3	Item #1
Common	Item #5	Form 9	Item #6	Form 9	Item #1
Form 1	Item #6	Form 2	Item #6		
Form 1	Item #7	Form 1	Item #7		
Form 2	Item #6	Form 10	Item #6		
Form 4	Item #6	Form 1	Item #6		
Form 4	Item #7	Form 3	Item #7		
Form 5	Item #6	Form 5	Item #6	Form 5	Item #1
Form 5	Item #7	Form 5	Item #7		
Form 6	Item #6	Form 6	Item #6		
Form 7	Item #6	Form 7	Item #6	Form 7	Item #1
Form 7	Item #7	Form 7	Item #7		
Form 8	Item #7	Form 8	Item #7		
Form 9	Item #6	Form 8	Item #6	Form 8	Item #1
Form 9	Item #7	Form 9	Item #7		
Form 10	Item #7	Form 10	Item #7		
Form 11	Item #6	Form 12	Item #6		
Form 11	Item #7	Form 6	Item #7		
Form 12	Item #7	Form 12	Item #7		
		Common	Item #1	Form 10	Item #1
		Common	Item #3	Form 1	Item #1
		Common	Item #4	Form 2	Item #1
		Common	Item #5	Form 11	Item #1
		Form 12	Item #6	Form 12	Item #1

LINKING ITEMS
GRADE 8
SCIENCE

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 1	Item #6	Form 4	Item #1
Common	Item #2	Form 4	Item #6	Form 10	Item #1
Common	Item #3	Form 10	Item #7		
Common	Item #4	Form 8	Item #6		
Common	Item #5	Form 3	Item #7		
Form 1	Item #6	Form 2	Item #6		
Form 2	Item #7	Form 4	Item #7		
Form 3	Item #6	Form 3	Item #6		
Form 3	Item #7	Form 8	Item #7	Form 3	Item #1
Form 5	Item #7	Form 5	Item #7		
Form 6	Item #7	Form 11	Item #7		
Form 7	Item #6	Form 7	Item #6		
Form 7	Item #7	Form 7	Item #7	Form 12	Item #1
Form 8	Item #6	Form 11	Item #6		
Form 8	Item #7	Form 1	Item #7		
Form 9	Item #6	Form 12	Item #6	Form 9	Item #1
Form 9	Item #7	Form 9	Item #7	Form 6	Item #1
Form 10	Item #6	Form 9	Item #6		
Form 11	Item #6	Form 10	Item #6		
Form 11	Item #7	Form 12	Item #7		
Form 12	Item #6	Form 6	Item #6		
		Common	Item #1	Form 5	Item #1
		Common	Item #2	Form 8	Item #1
		Common	Item #3	Form 2	Item #1
		Common	Item #4	Form 7	Item #1
		Common	Item #5	Form 1	Item #1
		Form 6	Item #7	Form 11	Item #1

LINKING ITEMS
GRADE 8
SOCIAL STUDIES

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 10	Item #6		
Common	Item #2	Form 9	Item #6		
Common	Item #3	Form 12	Item #7		
Common	Item #4	Form 11	Item #6		
Common	Item #5	Form 2	Item #6		
Form 1	Item #6	Form 12	Item #6		
Form 1	Item #7	Form 1	Item #7		
Form 2	Item #6	Form 1	Item #6	Form 1	Item #1
Form 2	Item #7	Form 2	Item #7		
Form 4	Item #6	Form 4	Item #6		
Form 4	Item #7	Form 4	Item #7	Form 4	Item #1
Form 5	Item #6	Form 5	Item #6		
Form 5	Item #7	Form 5	Item #7		
Form 7	Item #6	Form 7	Item #6	Form 7	Item #1
Form 7	Item #7	Form 8	Item #7		
Form 8	Item #6	Form 8	Item #6		
Form 8	Item #7	Form 7	Item #7		
Form 9	Item #7	Form 9	Item #7	Form 9	Item #1
Form 10	Item #6	Form 6	Item #6		
Form 10	Item #7	Form 10	Item #7		
Form 11	Item #7	Form 11	Item #7	Form 11	Item #1
Form 12	Item #7	Form 3	Item #6		
		Common	Item #1	Form 5	Item #1
		Common	Item #2	Form 10	Item #1
		Common	Item #3	Form 12	Item #1
		Common	Item #4	Form 8	Item #1
		Common	Item #5	Form 2	Item #1
		Form 3	Item #7	Form 3	Item #1
		Form 6	Item #7	Form 6	Item #1

LINKING ITEMS
GRADE 12
READING

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 1	Item #6		
Common	Item #2	Form 10	Item #7		
Common	Item #3	Form 7	Item #6		
Common	Item #4	Form 10	Item #6		
Common	Item #5	Form 3	Item #6	Form 3	Item #4
Form 2	Item #6	Form 2	Item #6	Form 2	Item #4
Form 2	Item #7	Form 2	Item #7		
Form 3	Item #6	Form 3	Item #7		
Form 4	Item #6	Form 4	Item #6	Form 4	Item #4
Form 4	Item #7	Form 4	Item #7		
Form 5	Item #6	Form 5	Item #6		
Form 5	Item #7	Form 5	Item #7	Form 5	Item #4
Form 8	Item #7	Form 8	Item #7		
Form 9	Item #6	Form 9	Item #6		
Form 9	Item #7	Form 9	Item #7		
Form 11	Item #7	Form 11	Item #7		
		Common	Item #1	Form 10	Item #4
		Common	Item #2	Form 1	Item #4
		Common	Item #3	Form 6	Item #4
		Common	Item #4	Form 11	Item #4
		Common	Item #5	Form 9	Item #4
		Form 12	Item #7	Form 12	Item #4
		Form 7	Item #7	Form 7	Item #4

LINKING ITEMS
GRADE 12
MATHEMATICS

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 2	Item #6		
Common	Item #3	Form 8	Item #6		
Common	Item #4	Form 2	Item #7		
Common	Item #5	Form 6	Item #6	Form 10	Item #1
Form 1	Item #7	Form 1	Item #7		
Form 2	Item #6	Form 7	Item #6		
Form 2	Item #7	Form 7	Item #7		
Form 3	Item #6	Form 3	Item #6		
Form 3	Item #7	Form 1	Item #6		
Form 4	Item #6	Form 4	Item #6		
Form 4	Item #7	Form 10	Item #7		
Form 5	Item #6	Form 5	Item #6		
Form 6	Item #6	Form 3	Item #7		
Form 6	Item #7	Form 6	Item #7		
Form 8	Item #7	Form 8	Item #7		
Form 9	Item #6	Form 9	Item #6	Form 2	Item #1
Form 11	Item #6	Form 11	Item #6		
Form 11	Item #7	Form 11	Item #7		
		Common	Item #1	Form 5	Item #1
		Common	Item #2	Form 4	Item #1
		Common	Item #3	Form 3	Item #1
		Common	Item #4	Form 6	Item #1
		Common	Item #5	Form 9	Item #1
		Form 11	Item #6	Form 7	Item #1

LINKING ITEMS
GRADE 12
SCIENCE

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 11	Item #6	Form 11	Item #1
Common	Item #2	Form 6	Item #6		
Common	Item #3	Form 5	Item #6	Form 4	Item #1
Common	Item #5	Form 11	Item #7		
Form 1	Item #6	Form 1	Item #6		
Form 1	Item #7	Form 1	Item #7		
Form 2	Item #6	Form 2	Item #6	Form 2	Item #1
Form 2	Item #7	Form 2	Item #7		
Form 3	Item #6	Form 3	Item #6	Form 3	Item #1
Form 3	Item #7	Form 3	Item #7		
Form 4	Item #6	Form 12	Item #7		
Form 4	Item #7	Form 9	Item #7		
Form 5	Item #6	Form 10	Item #6	Form 6	Item #1
Form 5	Item #7	Form 5	Item #7		
Form 7	Item #7	Form 7	Item #7		
Form 8	Item #6	Form 8	Item #6		
Form 8	Item #7	Form 8	Item #7		
Form 9	Item #6	Form 9	Item #6		
Form 9	Item #7	Fomr 4	Item #7		
Form 10	Item #6	Form 6	Item #7		
Form 10	Item #7	Form 10	Item #7		
Form 11	Item #6	Form 4	Item #6		
Form 12	Item #6	Form 12	Item #6	Form 8	Item #1
		Common	Item #1	Form 7	Item #1
		Common	Item #2	Form 1	Item #1
		Common	Item #3	Form 5	Item #1
		Common	Item #4	Form 10	Item #1
		Common	Item #5	Form 12	Item #1

LINKING ITEMS
GRADE 12
SOCIAL STUDIES

1993-94 LOCATION		1992-93 LOCATION		1991-92 LOCATION	
Common	Item #1	Form 1	Item #7		
Common	Item #2	Form 8	Item #6		
Common	Item #3	Form 1	Item #6	Form 1	Item #1
Common	Item #4	Form 3	Item #7		
Common	Item #5	Form 11	Item #6		
Form 1	Item #6	Form 10	Item #6	Form 10	Item #1
Form 1	Item #7	Form 10	Item #7		
Form 2	Item #6	Form 2	Item #6	Form 2	Item #1
Form 2	Item #7	Form 7	Item #7		
Form 3	Item #6	Form 3	Item #6		
Form 4	Item #6	Form 4	Item #6	Form 4	Item #1
Form 5	Item #7	Form 5	Item #7		
Form 6	Item #6	Form 6	Item #6		
Form 6	Item #7	Form 11	Item #7	Form 11	Item #1
Form 7	Item #6	Form 7	Item #6		
Form 8	Item #7	Form 8	Item #7		
Form 9	Item #6	Form 9	Item #6	Form 9	Item #1
Form 9	Item #7	Form 9	Item #7		
Form 11	Item #6	Form 5	Item #6		
Form 11	Item #7	Form 6	Item #7		
Form 12	Item #6	Form 12	Item #6		
Form 12	Item #7	Form 12	Item #7		
		Common	Item #1	Form 8	Item #1
		Common	Item #2	Form 6	Item #1
		Common	Item #3	Form 7	Item #1
		Common	Item #4	Form 12	Item #1
		Common	Item #5	Form 5	Item #1

Appendix 9-2
Open Response Test Raw Score to Performance Level Conversion Tables

1993-94 OPEN-RESPONSE TEST RAW SCORE TO PERFORMANCE LEVEL CONVERSION TABLE					
GRADE	CONTENT	PERFORMANCE LEVELS			
		NOVICE	APPRENTICE	PROFICIENT	DISTINGUISHED
4	READING	0-7	7-13	13-17	18-20
	MATH	0-8	8-14	15-16	16-20
	SCIENCE	0-9	9-16	16-17	18-20
	SOCIAL STUDIES	0-7	8-13	13-17	18-20
8	READING	0-7	7-14	14-18	18-20
	MATH	0-8	8-12	12-14	14-20
	SCIENCE	0-6	6-13	13-17	18-20
	SOCIAL STUDIES	0-7	7-12	12-16	17-20
12	READING	0-7	7-12	12-16	16-20
	MATH	0-4	4-8	8-12	12-20
	SCIENCE	0-5	5-12	13-16	16-20
	SOCIAL STUDIES	0-7	7-12	13-17	17-20

Appendix 9-3
Performance Event Raw Score to Performance Level Conversion Tables

1993-94 PERFORMANCE EVENT RAW SCORE TO PERFORMANCE LEVEL CONVERSION TABLE GRADE 4					
CONTENT	TASK	PERFORMANCE LEVELS			
		NOVICE	APPRENTICE	PROFICIENT	DISTINGUISHED
MATH	MA1	0-1	2-3	-	4
	MA3	0-2	3	-	4
	MA4	0-1	2	3	4
SOCIAL STUDIES	SS1	0-1	2-3	-	4
	SS2	0	1-2	3	4
	MA4	0-1	2-3	-	4
	SC3	0-1	2	3	4
SCIENCE	SC1	0	1-2	3	4
	SC2	0-1	2	3	4
	SC3	0-1	2-3	-	4
	MA3	0-1	2-3	-	4

1993-94 PERFORMANCE EVENT RAW SCORE TO PERFORMANCE LEVEL CONVERSION TABLE
GRADE 8

CONTENT	TASK	PERFORMANCE LEVELS			
		NOVICE	APPRENTICE	PROFICIENT	DISTINGUISHED
MATH	MA1	0-5	6	7-9	>9
	MA4	0-4	5-6	7	>7
SOCIAL STUDIES	SS1	0-5	6-8	9-10	>10
	SS2	0-3	4-7	8-10	>10
	MA4	0-2	3-4	5-7	>7
	SC3	1	2	3	>3
SCIENCE	SC1	0-9	10-12	13	>13
	SC2	0-6	7-12	13-14	>14
	SC3	0-3	4-6	7	>7

1993-94 PERFORMANCE EVENT RAW SCORE TO PERFORMANCE LEVEL CONVERSION TABLE GRADE 12					
CONTENT	TASK	PERFORMANCE LEVELS			
		NOVICE	APPRENTICE	PROFICIENT	DISTINGUISHED
MATH	MA1	0-2	3-6	7-11	>11
	MA2	0-13	14-16	17-20	>20
	MA3	0-2	3-6	7-13	>13
	MA4	0-4	5-6	7-13	>13
SOCIAL STUDIES	SS1	0-3	4-5	6	>6
	SS2	0-8	9-11	12-14	>14
	SS3	0-6	7-10	11-14	>14
	MA4	0-1	2-3	4-6	>6
	SC3	0-3	4-6	7-10	>10
SCIENCE	SC1	0-8	9-14	15-20	>20
	SC2	0-7	8-15	16-20	>20
	SC3	0-3	4-6	7-11	>11
	MA3	0-5	6-10	11-17	>17

Appendix 10-1
September 1994 School Report

KENTUCKY DEPARTMENT OF EDUCATION
CAPITAL PLAZA TOWER • 500 MERO STREET • FRANKFORT, KENTUCKY 40601
Thomas C. Boysen, Commissioner

Dear School Board Members and School Personnel:

The Kentucky Education Reform Act (KERA) of 1990 called for substantial changes in the educational system statewide. An innovative assessment program, the Kentucky Instructional Results Information System (KIRIS), was part of the Act. The assessment system is to be primarily performance-based by Spring 1996 and should drive instruction toward realization of goals specified in the Act.

During the 1993-94 school year, fourth, eighth, and twelfth grade students participated in several components of the assessment system: They completed written tests comprised of open-ended and multiple-choice questions in reading, mathematics, science, social studies, and an on-demand writing prompt (arts and humanities and practical living/vocational studies concepts were integrated into some of the open-ended questions and writing prompts); they created portfolios to demonstrate their skills in writing and mathematics; and they participated in performance events in mathematics, science, social studies, arts and humanities, and practical living/vocational studies. The assessments were developed under the direction of content area advisory committees of Kentucky educators who drafted, reviewed, and selected test questions. Items were also reviewed by a bias review committee.

I encourage you to review these results and use them in planning for school transformation. Please feel free to contact the staff of the Department of Education for assistance in interpreting and using the assessment results.

Sincerely,

Thomas C. Boysen

Thomas C. Boysen
Commissioner

KIRIS 1993-94
Assessment Curriculum Report

Code:

District:

School:

Grade: 12

September 1994

Contents of this Report

	Page
Introduction	2
Part I: Cognitive Results	
Overall Summary of Results	3
Reading Results	4
Mathematics Results	4
Science Results	5
Social Studies Results	5
Part II: Student Questionnaire Results	6
Part III: Subgroup and Selected Analyses	7
Part IV: District, State, and National Summary Data	10

Run Date: August 24, 1994

KIRIS 1993-94
Assessment Curriculum Report
September 1994

This report is part of a larger shipment of reports based on the 1993-94 Kentucky Instructional Results Information System (KIRIS) transitional testing results for students in grades 4, 8, and 12. Included in the shipment is a series of reports about the performance of individual students. This Student Assessment Curriculum Report summarizes information for the school. The report is divided into four parts: Part I details the cognitive results; Part II provides results to the student questionnaire items; Part III provides results for various student subgroups and some analyses cross-tabulating testing and questionnaire results; and Part IV provides summary data for the district, state, and nation.

Students in grades 4, 8, and 12 completed batteries of open-response and multiple-choice tests (referred to as transitional tests) in reading, mathematics, science, and social studies. To provide broad coverage of each subject domain, large numbers of test questions are developed and distributed across many different test forms; each student completed only one form. This approach, called "matrix sampling," produces highly reliable results at the building and district levels, while also reducing testing time for all students. The student-level results are based only on open-response questions **all** students completed ("common items") while this curriculum report is based on the entire battery of open-response questions. For reading, mathematics, science, and social studies, there were 29 open-response questions in each subject. Arts and humanities and practical living/vocational studies were included this year in the matrix portions of the tests, with open-response questions which were integrated with the other subject areas. Performance on the common and matrix items are weighted equally and averaged together to produce this report. Therefore, the number of students at each performance level may be reported as a fraction.

Schools are held accountable for all of the students enrolled in the school as of the first day of the testing window.

KERA states that "schools shall expect a high level of achievement of all students." It also states that "schools shall be rewarded for an increased proportion of successful students, including those students who are at risk of school failure." Therefore, there are virtually no exclusions from the testing. Students not participating in transitional tests were

- limited English-speaking students who have been enrolled in an English-speaking school for fewer than two years;
- foreign exchange students;
- students with moderate to severe disabilities who complete an alternate portfolio;
- students with severe medical conditions such that participation would threaten health or well-being.

The number and percent of students who did not participate for these reasons are provided on the bottom of page 3. Any other student for whom the school is accountable but who was not tested is assigned to the "Novice" performance level. The number and percentage of students not tested in transitional testing, for whom a "Novice" rating is reported in the subject area reports, are also provided on page 3.

KIRIS 1993-94
Assessment Curriculum Report
September 1994

Code:
District:
School:
Grade: 12

PART I: Cognitive Results

Overall Summary of Results

	Performance Level									
	Distinguished		Proficient		Apprentice		Novice		Non-tested	
	#	%	#	%	#	%	#	%	#	%
Reading	0.0	0	3.5	4	54.5	66	25.0	30	0.0	0
Mathematics	2.5	3	5.5	7	41.5	50	33.5	40	0.0	0
Science	0.5	1	1.0	1	59.5	72	21.0	25	1.0	1
Social Studies	0.0	0	6.0	7	37.5	45	37.5	45	2.0	2
Arts and Humanities		0		2		9		88		2
Practical Living/Vocational Studies		0		5		12		83		0
Writing Portfolio		Data will not be released until all portfolio reviews are completed.								

NOTE: The results in reading, mathematics, science, and social studies are based on the 29 open-response (common and matrix) questions in each subject area. Results for arts and humanities and practical living/vocational studies are based on matrix open-response questions which were integrated with the other subject areas.

NOTE: Since performance on the common and matrix items is weighted equally and averaged together to produce this report, the number of students at each performance level may be reported as a fraction.

Students not participating:	#	%
Eligible for Alternate Portfolio	0	0
Limited-English speakers	0	0
Medical exemption	0	0

325

KIRIS 1993-94
Assessment Curriculum Report
September 1994

Code:
District:
School:
Grade: 12

Reading Results

	Distinguished		Proficient		Apprentice		Novice *	
	#	%	#	%	#	%	#	%
1993-94	0.0	0	3.5	4	54.5	66	25.0	30
1992-93		0		5		38		56
1991-92		0		0		0		0

Last year you had 5% Proficient or Distinguished students; this year you had 4% Proficient or Distinguished.

	Distinguished	Proficient	Apprentice	Novice *
PASSAGE TYPE	%	%	%	%
Literary	0	4	66	30
Informational	0	7	80	13
Practical/Workplace	0	13	55	31
Persuasive	0	1	86	13

You had the highest percentage of Proficient or Distinguished students in Practical/Workplace and the lowest percentage of Proficient or Distinguished in Persuasive.

Mathematics Results

	Distinguished		Proficient		Apprentice		Novice *	
	#	%	#	%	#	%	#	%
1993-94	2.5	3	5.5	7	41.5	50	33.5	40
1992-93		3		7		38		52
1991-92		1		8		49		43

Last year you had 10% Proficient or Distinguished students; this year you had 10% Proficient or Distinguished.

	Distinguished	Proficient	Apprentice	Novice *
REPORTING CATEGORY	%	%	%	%
Number, Procedures, and Structure	2	6	64	28
Space & Dimensionality, and Measurement	6	16	58	20
Change and Data	1	5	64	30

You had the highest percentage of Proficient or Distinguished students in Space & Dimensionality, and Measurement and the lowest percentage of Proficient or Distinguished in Change and Data.

---

*Any non-tested students have been assigned to the Novice category.

KIRIS 1993-94
Assessment Curriculum Report
September 1994

Code:
District:
School:
Grade: 12

Science Results

	Distinguished		Proficient		Apprentice		Novice *	
	#	%	#	%	#	%	#	%
1993-94	0.5	1	1.0	1	59.5	72	22.0	27
1992-93		0		3		51		46
1991-92		0		0		0		0

Last year you had 3% Proficient or Distinguished students; this year you had 2% Proficient or Distinguished.

	Distinguished	Proficient	Apprentice	Novice *
REPORTING CATEGORY	%	%	%	%
Nature of Scientific Activity, Models & Scales	17	81	2	0
Patterns, Constancy, and Evolution	13	83	2	1
Systems	16	77	6	1

You had the highest percentage of Proficient or Distinguished students in Systems and the lowest percentage of Proficient or Distinguished in Nature of Scientific Activity, Models & Scales.

Social Studies Results

	Distinguished		Proficient		Apprentice		Novice *	
	#	%	#	%	#	%	#	%
1993-94	0.0	0	6.0	7	37.5	45	39.5	48
1992-93		0		3		34		63
1991-92		1		6		44		49

Last year you had 3% Proficient or Distinguished students; this year you had 7% Proficient or Distinguished.

	Distinguished	Proficient	Apprentice	Novice *
REPORTING CATEGORY	%	%	%	%
Democratic Principles and Political Systems	30	49	20	0
Social Systems, Economic Systems, and Cultural Diversity	41	48	10	1
Relation of Geography to Human Activity and Historical Perspective	48	43	8	0
Social Studies & Inquiry Skills	57	33	10	0

You had the highest percentage of Proficient or Distinguished students in Democratic Principles and Political Systems and the lowest percentage of Proficient or Distinguished in Relation of Geography \ Historical Perspective.

---

*Any non-tested students have been assigned to the Novice category.

PART II: Student Questionnaire Results

	#	%
What is the highest education level reached by your parents? (mark only one answer - the level for your parent who had the most education)		
did not finish high school	17	20
graduated from high school	21	25
had some education after high school (for example, business or technical school)	21	25
graduated from college	9	11
advanced degree	10	12
How much time do you usually spend on homework each day?		
less than 1 hour	43	52
1 hour	21	25
between 1 and 2 hours	12	14
between 2 and 3 hours	2	2
more than 3 hours	2	2
How many hours per week do you usually work in a part-time job?		
none	42	51
10 or fewer	12	14
11 to 15	7	8
16 to 20	7	8
more than 20	11	13
How many days of school did you miss in January?		
none	40	48
1 or 2 days	24	29
3 or 4 days	7	8
5 to 8 days	2	2
more than 8 days	6	7
What is the one thing that is likely to take the largest share of your time in the year after you leave high school?		
working full-time	20	24
attending a two-year college or a vocational, technical, or business school	15	18
attending a four-year college	39	47
enlisting in the military	4	5
Which best describes your use of a calculator on the mathematics test you took a few days ago?		
I did not use a calculator.	20	24
I used my own calculator.	27	33
I used a calculator provided by the school.	32	39
Which best describes your use of a calculator on the science test you took a few days ago?		
I did not use a calculator.	59	71
I used my own calculator.	7	8
I used a calculator provided by the school.	14	17
Which of the following best describes your grades so far in school?		
mostly "A"s		
mostly "B"s		
a mix of "B"s and "C"s		
mostly "C"s		
mostly "D"s or lower		

	#	%
If you are currently taking a course in English, what is the level of the course?		
advanced placement or honors	6	7
college preparatory	24	29
regular or non-college preparatory	42	51
courses in my school are not leveled	6	7
I am not currently taking an English course	1	1
What is the mathematics course you are currently taking? (If you are not currently taking mathematics, what is the last mathematics course you took?)		
General Math/Business Math	26	31
Geometry or Algebra	37	45
Advanced Algebra	4	5
Pre-Calculus or Calculus	12	14

By the time you graduate from high school, how many courses will you have completed in each subject area while you were in high school?

	0		1		2		3		4 or more	
	#	%	#	%	#	%	#	%	#	%
Eng./Lang. Arts	1	1	0	0	1	1	0	0	78	94
Mathematics	2	2	0	0	1	1	60	72	17	20
Science	1	1	1	1	36	43	16	19	26	31
Social Studies	1	1	3	4	59	71	10	12	7	8
Foreign Language	43	52	12	14	19	23	3	4	3	4

How often do you participate in the following activities?

	2 or more times/week		about once a week		about once a month		never	
	#	%	#	%	#	%	#	%
working in groups	54	65	16	19	7	8	3	4
working on projects	15	18	26	31	31	37	8	10
looking up information	31	37	30	36	16	19	3	4
giving reports out loud	0	0	5	6	55	66	19	23
using a calculator	28	34	13	16	16	19	18	22
reading books students choose	0	0	6	7	23	28	42	51
writing (journals, stories, other)	67	81	5	6	3	4	5	6
solving mathematics problems	33	40	12	14	12	14	19	23
using mathematics manipulatives	21	25	19	23	6	7	27	33

How many of the questions on the test covered material that you learned in school in this or any grade?

	most of the questions		some of the questions		none or very few	
	#	%	#	%	#	%
atics	26	31	42	51	11	13
	17	20	55	66	8	10
tudies	22	27	33	40	24	29
Arts and Humanities	11	13	27	33	40	48
Practical living or Vocational studies	15	18	30	36	31	37

BEST COPY AVAILABLE

Code:
Grade: 12

PART III: Subgroup and Selected Analyses

	Male								Female							
	Distinguished		Proficient		Apprentice		Novice		Distinguished		Proficient		Apprentice		Novice	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Reading	0.0	0	1.0	2	28.0	60	18.0	38	0.0	0	2.5	7	26.5	74	7.0	19
Mathematics	2.5	5	2.5	5	20.5	44	21.5	46	0.0	0	3.0	8	21.0	58	12.0	33
Science	0.5	1	0.5	1	31.0	67	14.0	30	0.0	0	0.5	1	28.5	79	7.0	19
Social Studies	0.0	0	4.0	9	17.5	38	24.5	53	0.0	0	2.0	6	20.0	57	13.0	37

	White								Non-white							
	Distinguished		Proficient		Apprentice		Novice		Distinguished		Proficient		Apprentice		Novice	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Reading	0.0	0	3.0	4	52.0	66	24.0	30								
Mathematics	2.5	3	5.5	7	39.0	49	32.0	41								
Science	0.5	1	0.5	1	56.5	72	20.5	26								
Social Studies	0.0	0	6.0	8	34.0	44	37.0	48								

	Chapter 1								Non-chapter 1							
	Distinguished		Proficient		Apprentice		Novice		Distinguished		Proficient		Apprentice		Novice	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Reading	0.0	0	1.0	6	5.0	31	10.0	63	0.0	0	2.5	4	46.5	74	14.0	22
Mathematics	0.0	0	0.0	0	7.0	44	9.0	56	2.0	3	5.5	9	33.0	52	22.5	36
Science	0.0	0	0.0	0	8.0	53	7.0	47	0.5	1	1.0	2	49.0	78	12.5	20
Social Studies	0.0	0	0.0	0	4.0	25	12.0	75	0.0	0	6.0	10	31.5	52	23.5	39

Note: To protect anonymity, no data are reported if a category includes fewer than 10 students. Also, these analyses are based on tested students.

Run Date: August 24, 1994

Performance Level by Student Questionnaire Responses

Time Spent on Homework Each Day	% Proficient or Distinguished			
	Reading	Mathematics	Science	Social Studies
Less than 1 hour	6	6	2	7
1 hour	5	14	0	8
Between 1 and 2 hours	0	21	4	13
2 hours or more				

Hours Per Week Worked in a Part-time Job	% Proficient or Distinguished			
	Reading	Mathematics	Science	Social Studies
None - 10 hours	6	12	2	8
11-20 hours or more	2	2	0	10

Post-High School Plans	% Proficient or Distinguished			
	Reading	Mathematics	Science	Social Studies
Working	0	0	0	3
Attending 2- or 4-year college/school	7	14	3	10
Military				

Parent Education Level	% Proficient or Distinguished			
	Reading	Mathematics	Science	Social Studies
Did not finish high school	3	0	0	8
Graduated from high school	3	7	0	3
Education after high school	5	14	4	11

Note: To [illegible]ct anonymity, no data are reported if a category includes fewer t[illegible]0 students. Also, these analyses are based on tested students.

Run Dat[illegible]ust 24, 1994

Code:
District:
School:
Grade: 12

Performance Level by Student Questionnaire Responses (cont.)

	Reading		Mathematics		Science		Social Studies	
	% Proficient or Distinguished		% Proficient or Distinguished		% Proficient or Distinguished		% Proficient or Distinguished	
Frequency of Activity as Reported by Students	1/week or more	1/month or less	1/week or more	1/month or less	1/week or more	1/month or less	1/week or more	1/month or less
working in groups	5	5	9	20	2	0	8	5
working on projects	5	4	11	9	2	2	6	9
looking up information	3	8	9	14	3	0	8	8
giving reports outs loud		4		11		2		8
reading books students choose		4		12		1		6
writing (journals, stories, other)	5		11		2		9	
solving mathematics problems	N/A	N/A	16	3	3	0	N/A	N/A
using a calculator	N/A	N/A	17	3	4	0	N/A	N/A
using mathematics manipulatives	N/A	N/A	17	5	4	0	N/A	N/A

Note: To protect anonymity, no data are reported if a category includes fewer than 10 students. Also, these analyses are based on tested students.

KIRIS 1993-94
Assessment Curriculum Report
September 1994

Code:
District:
School:
Grade: 12

PART IV: District, State, and National Summary Data

	Performance Level	Reading	Mathematics	Science	Social Studies
Percentage of Students in District at Each Performance Level	Distinguished	0	3	1	0
	Proficient	4	7	1	7
	Apprentice	66	50	72	45
	Novice	30	40	27	48
Percentage of Students in State at Each Performance Level	Distinguished	2	6	1	2
	Proficient	14	13	6	16
	Apprentice	55	48	69	49
	Novice	29	33	24	34
Percentage of Students Nationwide at Each Performance Level	Distinguished	2	5	N/A	N/A
	Proficient	12	12	N/A	N/A
	Apprentice	49	42	N/A	N/A
	Novice	37	41	N/A	N/A

1993-94

Kentucky Instructional Results Information System

Grade 4 ITEM LEVEL REPORT FOR PERFORMANCE-BASED TASKS

(Open-ended items and writing on-demand)

Code:
District:
School:

Page: 1 of 1
Group Size: 25
September, 1994

	Test Item #	COMMON Reading 1 2 3 4 5	PL	Math 1 2 3 4 5	PL	Science 1 2 3 4 5	PL	Social Studies 1 2 3 4 5	PL	MATRIX SAMPLED Form	Reading 1 2	Math 1 2	Science 1 2	Social Studies 1 2	A & H 1 2	PL / VS 1 2	WRITING (On-Demand) Score	Task #
I	0440847	3 2 3 2 2	A	2 4 3 1 3	A	1 4 3 2 2	A	2 2 1 2 0	N	11	3 1	4 0	2 2	2 1	1 X	2 X	1	21
B.	0440823	2 2 2 2 2	A	2 4 2 2 3	A	3 2 2 3 1	A	2 3 2 1 1	A	03	0 0	2 1	0 2	2 1	0 X	X X	1	05
Bl	0440846	2 4 4 3 3	P	3 2 4 4 3	P	2 4 3 2 3	A	4 4 3 3 4	D	05	4 4	3 2	4 3	4 4	4 X	3 X	3	10
B	0440824	2 4 4 2 3	P	3 4 2 2 1	A	1 2 1 2 1	N	1 2 2 3 0	A	12	2 3	1 0	2 3	1 1	2 X	1 X	1	23
CC	0440833	3 3 3 3 3	P	4 3 4 4 3	D	4 4 4 2 4	D	3 4 4 2 4	P	01	4 1	2 2	3 1	3 4	X X	4 X	1	01
D	0440831	3 4 4 3 4	D	3 4 4 3 3	D	4 3 3 2 2	A	3 4 4 4 3	D	05	2 3	2 3	3 4	4 4	4 X	4 X	2	10
[illegible]	0440849	3 4 3 3 4	P	4 2 4 4 3	D	4 4 4 4 4	D	2 4 4 4 4	D	02	2 2	4 3	4 3	3 4	X X	3 X	3	04
[illegible]	0440836	2 3 3 3 3	P	1 3 3 4 3	A	4 4 3 4 3	D	3 4 3 3 4	P	01	4 2	2 3	3 2	4 4	X X	4 X	2	01
[illegible]	0440840	2 2 3 3 3	A	2 3 4 4 3	P	3 3 3 2 2	A	3 4 3 3 4	P	08	1 4	4 2	3 4	4 3	4 X	2 X	1	16
GA	0440848	2 3 3 2 3	P	3 3 2 4 3	P	2 3 3 2 3	A	3 4 3 3 4	P	04	1 3	3 3	1 1	4 4	3 X	1 X	1	08
[illegible]	0440838	2 4 4 3 3	P	3 2 3 4 3	P	3 4 4 2 3	P	3 4 4 3 4	D	07	2 2	2 3	4 4	2 4	3 X	4 X	1	14
[illegible]	0440841	3 3 4 3 3	P	2 3 3 4 3	P	4 4 2 3 3	A	4 4 4 2 4	D	10	3 3	4 3	2 2	2 4	4 X	4 X	1	19
H	0440835	3 3 4 3 4	P	0 3 4 2 3	A	3 4 2 2 4	A	2 4 4 4 4	D	11	3 2	4 4	3 3	4 4	4 X	2 X	2	21
J	0440839	3 4 4 3 4	D	4 2 4 3 3	P	3 4 2 2 3	A	3 4 3 4 4	D	06	3 3	4 4	4 2	4 4	2 X	3 X	3	12
JO	0440825	3 3 3 2 3	P	4 3 4 3 4	D	2 3 2 2 1	A	1 2 1 2 1	N	02	2 3	4 3	0 3	2 1	X X	2 X	1	03
LI	0440842	2 2 4 4 4	P	3 3 4 4 3	D	3 4 3 2 3	A	3 3 3 4 2	P	11	3 3	4 4	2 3	4 3	3 X	3 X	1	21
M.	0440845	3 3 3 2 2	P	4 4 4 4 3	D	4 4 2 4 3	P	2 4 4 3 4	P	12	3 4	2 1	4 3	4 4	3 X	4 X	1	23
M	0440832	3 2 4 2 4	P	2 3 2 3 3	A	4 4 4 2 2	A	4 4 4 4 4	D	12	2 4	4 3	3 2	3 4	2 X	4 X	2	23
M E	0440827	2 1 0 0 2	N	2 2 3 0 B	N	0 0 0 0 0	N	1 1 0 1 0	N	10	0 0	B B	0 0	0 2	2 X	B X	1	19
P R	0440826	2 1 2 1 2	A	3 4 3 1 3	A	2 4 1 2 3	A	2 1 3 2 2	A	01	1 1	0 2	1 1	2 1	X X	1 X	1	01
S	0440843	3 3 4 2 4	P	4 4 4 4 3	D	3 4 2 4 4	P	2 4 3 3 3	P	09	3 2	4 4	4 2	4 4	X X	3 X	2	18
S.	0440829	3 3 B 2 2	A	2 0 4 1 4	A	1 1 3 2 3	A	1 1 0 2 0	N	06	1 1	1 1	1 1	1 2	1 X	1 X	1	12
S	0440830	3 2 1 2 2	A	4 3 3 1 4	P	1 1 1 3 3	N	1 1 0 2 3	N	07	2 1	4 2	1 1	2 1	2 X	1 X	1	14
S	0440822	2 2 2 2 2	A	2 3 2 2 3	A	1 3 0 1 1	N	2 2 2 0 3	A	04	2 2	3 2	1 0	2 2	2 X	1 X	1	08
S	0440834	3 3 4 3 4	P	4 2 3 3 3	P	3 4 3 3 3	P	2 4 4 4 3	P	03	4 3	4 4	4 3	2 4	4 X	X X	3	05
School Mean Score		2.6 2.8 3.0 2.4 3.0		2.8 2.9 3.3 2.8 2.9		2.6 3.2 2.4 2.4 2.6		2.4 3.1 2.7 2.7 2.8										
District Mean Score		2.3 2.3 2.5 2.2 2.9		2.5 2.5 2.6 2.5 2.7		2.0 3.1 2.2 2.5 2.5		2.2 2.7 2.5 2.4 2.2										
State Mean Score		1.7 1.7 1.9 1.7 2.4		1.9 1.9 1.6 1.8 1.4		1.4 2.6 1.6 2.2 1.9		1.8 2.1 1.5 1.8 1.6										

* Reassigned from non-A1 school.

NR = No response
OT = Off topic
NE = Not English

LEGEND: B = NO ATTEMPT BLANK = NO AVAILABLE DATA X = NO ITEM IN TEST FORM PL = PERFORMANCE LEVEL (Novice, Apprentice, Proficient, Distinguished)

Run Date: August 23, 1994

BEST COPY AVAILABLE

1993-94

Kentucky Instructional Results Information System

Grade 4

ITEM LEVEL REPORT FOR MULTIPLE CHOICE

Code:
District:
School:

Page: 1 of 1
Group Size: 25
September, 1994

	Test Item #	Reading			Mathematics		Science		Social Studies	
		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
LA	0440847	+++++CC+BB	A++A+CB++A	+CA+D	++D+BD+AD+	C+C+A+AD++	BD+AD++++D	AD+A+D+CD+	A+++B++++C	DAB++ADB+D
	0440823	++++CB+++A	C++++CB+AA	++++B	A+++++++++C	A++D+++++A	++++A+++++	+C++A+++++	+++++++BD+	+D++AC+++B
BL ER N	0440846	+++++C+A+A	A+++DC++AB	++++C	D+++BD+A++	++++A+A++B	++++D++C++	++++AD++D+	+++++++++A	+AB+++++++
	0440824	++B+CCA++C	C++A+D++BA	+++AB	+CC++++++C	C+C+A++CDB	+++ABA+D++	+++BD++ADA	+AD+++++A+	D+A+++DB++
	0440833	+++++C++AA	++C++C+++A	++++D	+++++++C++	+++++D+C+A	++++D++CB+	++B+A++ACC	+++++++++A	++B+++D+++
T IA	0440831	+++++C+++A	C++++A+++A	D+++B	+++++++++C	+C+++++C++	++++D+++C+	++++D+++B+	++++++++++	++++++++++
D	0440849	+++++C+++A	+++++C++AA	++++B	++++++++++	+++++C+++A	+++++++A++	++B+++++++	++++++++++	++++++++++
EV	0440836	+++++C+++A	+++++CA++A	++++B	+++++++++C	+D++++++++	++++++++++	++++++++++	++++++++++	++++++++++
F Y	0440840	+++++C+++A	+++++C+++A	++++B	B+D++++++C	+++D+D+D+A	+++++++++D	+A++AD++DB	++++++++++	++++++++++
C DA	0440848	+++++A+++C	A+CADCB+CA	D+++C	D++CBBC+++	+++BAC+C++	++++A++C++	++++A++B++	++++A+++++	++C+++++A+
G T	0440838	+++++C++BA	+++++C+++A	++CAD	+++++D+++C	+++++++++C	+++++++D++	++++A+++C+	A+++++++++	++++++++++
IR A	0440841	+++++C+++A	+++++C+++A	+++CD	+++++++A+C	+++++A+C++	+++++++C++	++++A++BC+	++++++++++	++++++++++
	0440835	+++++C+++B	+++++C+++A	++++B	++++++++++	++++++++++	+++++++A++	++++++++++	++++++++++	++++++++++
	0440839	+++++C+++A	A++++C+++A	++++B	+++++++++C	+B++++++++	++++++++++	++++++++++	++++++++++	++++++++++
An	0440825	+++ACCAC+B	+AD+BB+++B	BCDBB	++A+B+A+CC	+BB++D+A++	+++DACDDCA	B++++++BD+	+++CB+AD++	A++D+B+CDD
L	0440842	+++++C+++A	+++++C+++B	+C++B	+++++++++C	+D+++D+CB+	+++++++D++	+++BA++BD+	++++++++++	+++++++B++
M EL	0440845	+++++C+++A	+++++C+++A	+C++B	+++++++++C	+++++++++C	++++DA+C+A	+A++++C+B+	++++++++A+	++++++++D+
	0440832	+++++C+++A	+++++C+++A	++++B	+++++++A+C	++++++++++	+++D+++C++	+++B++++B+	++++++++++	+++++++++0
	0440827	+++ACD++CB	AC+DBCAB+A	+DCAC	B+AAACCD+C	C+CDA++ABA	++BDDCACAB	BAD+CC+BC+	++CAC+++CD	DBCA+A+DB+
POK	0440826	+D+++BA+AA	+B+++C+++B	++++D	++D++C++++	++++ADAC+C	++++B++++A	DA+B++A+++	++D+++C+++	DD++0C+BBB
SI	0440843	+++++C+++A	A++++C+++B	++++D	+++++++A++	+D+++C++B+	++++A++D++	+++A++++++	++++++A+++	+++++++B++
S	0440829	++BD+BBD+A	CB+ACDBB+A	+C++B	D+C++B+ABC	A+C++C+DDB	++++B++C++	B+C+ABA+D+	+++++++B+A	D+A+AB+CB+
S	0440830	+ADA+B+C+A	AC++CADCAB	BC+CD	A++++DC++C	++++BC++++	+B+D+D+A+B	D+CB+B+CB+	++++C+DBDA	ADC+ACABDC
S A	0440822	+D+CBAA++B	CB+++D+++A	DC++D	ACC++++ACC	++C+AD++++	+++AD+CCAD	+A+CABACCA	++++CB+++A	+DC++BC+B+
SM	0440834	+++++C+++A	+++++C+++A	++++B	++++++++++	++++++A+++	+++++++A++	++++++++++	++++++++++	++++++++++
School Percentage Correct		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
District Percentage Correct		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
State Percentage Correct		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

* Reassigned from non-A1 school.

** Reader awareness questions - all scored as correct.

LEGEND: + = CORRECT ANSWER A,B,C, or D = INCORRECT OPTION CHOSEN 0 = NO ATTEMPT BLANK = NO DATA AVAILABLE

Run Date: August 23, 1994

BEST COPY AVAILABLE

1993-94

Kentucky Instructional Results Information System

Grade 4 STUDENT LISTING

Code:
District:
School:

Page: 1 of 1
Group Size: 25
September, 1994

			Performance Levels				Total Battery Percentile Rank	Kentucky NCEs			
			Reading	Math	Science	Social Studies		Reading	Math	Science	Social Studies
[illegible]	[illegible]	0440847	A	A	A	N	71	65	66	56	41
[illegible]	[illegible]	0440823	A	A	A	A	68	63	74	74	60
[illegible]	[illegible]	0440846	P	P	A	D	99	93	81	87	99
[illegible]	[illegible]	0440824	P	A	N	A	65	80	64	36	53
[illegible]	[illegible]	0440833	P	D	D	P	99	93	99	99	99
[illegible]	[illegible]	0440831	D	D	A	D	99	99	99	90	99
[illegible]	[illegible]	0440849	P	D	D	D	99	99	99	99	99
[illegible]	[illegible]	0440836	P	A	D	P	99	90	85	99	99
[illegible]	[illegible]	0440840	A	P	A	P	97	87	81	77	99
[illegible]	[illegible]	0440848	P	P	A	P	96	67	74	83	99
[illegible]	[illegible]	0440838	P	P	P	D	99	93	88	99	99
[illegible]	[illegible]	0440841	P	P	A	D	99	99	85	99	99
[illegible]	[illegible]	0440835	P	A	A	D	99	99	80	99	99
[illegible]	[illegible]	0440839	D	P	A	D	99	99	93	99	99
[illegible]	[illegible]	0440825	P	D	A	N	83	65	88	48	41
[illegible]	[illegible]	0440842	P	D	A	P	99	99	93	90	99
[illegible]	[illegible]	0440845	P	D	P	P	99	85	99	99	99
[illegible]	[illegible]	0440832	P	A	A	D	99	99	80	99	99
[illegible]	[illegible]	0440827	N	N	N	N	4	26	38	1	17
[illegible]	[illegible]	0440826	A	A	A	A	71	53	74	71	58
[illegible]	[illegible]	0440843	P	D	P	P	99	99	99	99	99
[illegible]	[illegible]	0440829	A	A	A	N	44	47	56	54	32
[illegible]	[illegible]	0440830	A	P	N	N	62	43	80	41	32
[illegible]	[illegible]	0440822	A	A	N	A	50	53	64	20	55
[illegible]	[illegible]	0440834	P	P	P	P	99	99	93	99	99

3ER OF EXEMPTIONS
O ESL
O ALTERNATE PORTFOLIO
O MEDICAL EXCLUSION
O FOREIGN EXCHANGE

Summation of these data will not match summary data on the curriculum report since this list is based only on common items while the curriculum report is based on a combination of common and matrix items.

Run Date: August 23, 1994

** Total Battery percentile rank based on open-ended questions only

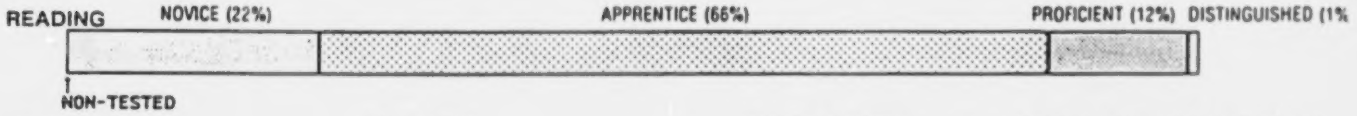
N = Novice, A = Apprentice, P = Proficient, D = Distinguished

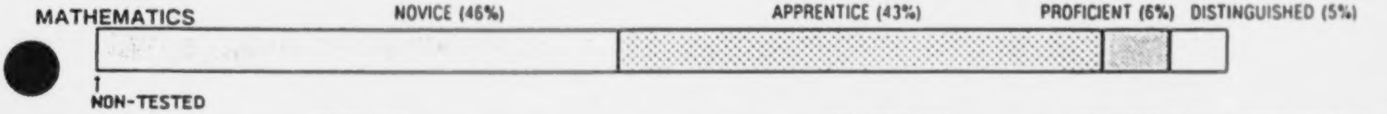
BEST COPY AVAILABLE

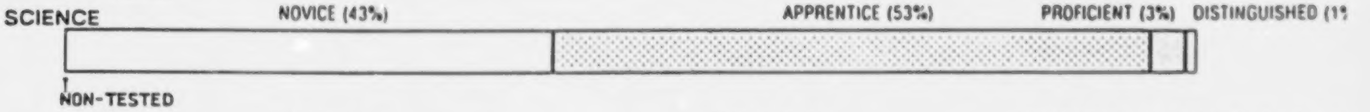
INDIVIDUAL STUDENT REPORT

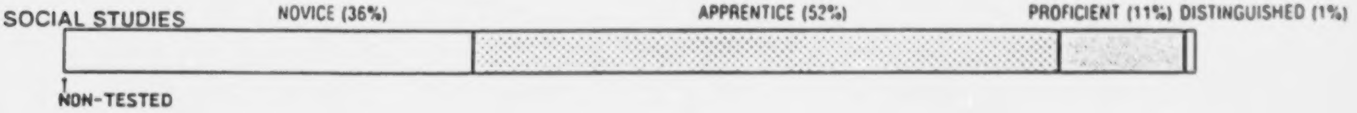
Report To: and Parents/Guardians
Grade: 4
School:
District:

This report is to inform you of the results of the test you took in 1993-94 through the Kentucky Instructional Results Information System (KIRIS). Results are described in terms of performance levels. KERA, the education reform legislation, called for all students to achieve at high levels. That high level achievement is defined as the Proficient level. The bar shows the percentage of Kentucky's fourth graders who scored in each category, with an arrow pointing to your score.









You were not tested in 1993-94.

Writing Portfolio results will not be released until all portfolio reviews are completed. National comparative data for science and social studies are not available. Details about the performance levels are available from your school. Should you have other questions, you are encouraged to contact your school staff.

Produced for the Kentucky Department of Education, September 1994

Appendix 10-2
January 1995 School Report

KIRIS BIENNIUM I
SCHOOL ACCOUNTABILITY REPORT

Code:
District:
School:
Grade:

Cognitive Areas

	Performance Level											
	Distinguished %			Proficient %			Apprentice %			Novice %		
	1991-92	1992-93	1993-94	1991-92	1992-93	1993-94	1991-92	1992-93	1993-94	1991-92	1992-93	1993-94
Reading												
Transitional Total	0.58	0.48	0.49	2.91	1.92	12.62	50.00	42.79	51.94	46.51	54.81	34.95
Common	0.00	0.96	0.97	1.16	2.88	14.56	52.33	38.46	47.57	46.51	57.69	36.89
Matrix	1.16	0.00	0.00	4.65	0.96	10.68	47.67	47.12	56.31	46.51	51.92	33.01
Combined	0.5800	0.4800	0.4850	2.9050	1.9200	12.6200	50.0000	42.7900	51.9400	46.5100	54.8050	34.9500
Mathematics												
Transitional Total	2.91	2.89	6.31	8.72	7.69	6.32	48.84	34.14	51.46	39.54	55.29	35.92
Common	3.49	5.77	8.74	5.81	8.65	6.80	50.00	34.62	46.60	40.70	50.96	37.86
Matrix	2.33	0.00	3.88	11.63	6.73	5.83	47.67	33.65	56.31	38.37	59.62	33.98
Performance Events	0.00	0.00	4.00	3.23	3.85	20.00	48.39	84.62	44.00	48.39	11.54	32.00
Combined	2.3280	2.3080	5.8500	7.6220	6.9220	9.0500	48.7460	44.2320	49.9600	41.3060	46.5400	35.1400
Science												
Transitional Total	1.17	0.00	0.97	2.91	2.89	4.37	57.56	52.89	68.45	38.37	44.23	26.22
Common	0.00	0.00	1.94	2.33	3.85	4.85	60.47	50.00	65.05	37.21	46.15	28.16
Matrix	2.33	0.00	0.00	3.49	1.92	3.88	54.65	55.77	71.84	39.53	42.31	24.27
Performance Events	0.00	5.00	0.00	0.00	15.00	4.17	63.16	50.00	37.50	36.84	30.00	58.33
Combined	0.9320	1.0000	0.7800	2.3280	5.3080	4.3300	58.6800	52.3080	62.2600	38.0640	41.3840	32.6400
Social Studies												
Transitional Total	1.16	0.48	1.94	6.98	3.37	8.74	44.19	36.54	53.88	47.68	59.62	35.44
Common	1.16	0.00	0.97	8.14	1.92	7.77	44.19	42.31	58.25	46.51	55.77	33.01
Matrix	1.16	0.96	2.91	5.81	4.81	9.71	44.19	30.77	49.51	48.84	63.46	37.86
Performance Events	0.00	0.00	0.00	5.71	3.70	4.00	60.00	66.67	28.00	34.29	29.63	68.00
Combined	0.9280	0.3840	1.5500	6.7220	3.4320	7.7900	47.3520	42.5660	48.7000	44.9980	53.6180	41.9500
Writing												
Portfolio	0.00	0.00	2.91	17.44	10.58	17.48	32.56	42.31	33.01	50.00	47.12	46.60

KIRIS Biennium 1
School Accountability Report

Code
District
School
Grade

Cognitive Index

	1991-92	1992-93	1993-94
Reading	23.7170	19.7080	34.0750
Mathematics	30.3796	27.8460	37.2248
Science	27.1048	27.6312	30.3148
Social Studies	26.9620	20.9960	29.4464
Writing	30.4640	27.5040	34.7580

Noncognitive Areas

	1991-92 %	1992-93 %	1993-94 %
Attendance	94.43	93.95	94.39
Drop-out	1.78	3.88	8.19
Retention	2.27	4.67	5.61
Successful Transition	86.08	90.91	97.03
Noncognitive Index	92.8850	93.6928	94.4126

Year Data Are Collected			Summary Data	
1991-92	1992-93	1993-94		
			Baseline Accountability Index for 1992-94 Biennium	38.6
			Threshold for 1992-94 Biennium	44.7
			Accountability Index for 1992-93 School Year	36.2
			Accountability Index for 1993-94 School Year	43.4
			Accountability Index for 1992-94 Biennium	39.8

Your final Biennium 1 accountability index, which is the weighted average of your 1992-1993 and 1993-1994 accountability indices, is greater than or equal to your baseline, but less than your threshold. Therefore, your school is required to develop a school improvement plan and is eligible for school improvement funds to assist with implementation of your plan.

Please consult the enclosed Program Review on Rewards and Assistance Implementation, dated December 2, 1994, for information about specific next steps you must take.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030642

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (9/97)